(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
31 August 2023 (31.08.2023)

WIPO | PCT

(10) International Publication Number
WO 2023/161266 A1

(51) International Patent Classification:
C12N 9/00 (2006.01) *C12N 15/52* (2006.01)
A61K 38/53 (2006.01)

(21) International Application Number:
PCT/EP2023/054401

(22) International Filing Date:
22 February 2023 (22.02.2023)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
22157958.4 22 February 2022 (22.02.2022) EP

(71) Applicants: **UNIVERSITÉ LIBRE DE BRUXELLES** [BE/BE]; Avenue Franklin D. Roosevelt 50, Bruxelles, 1050 (BE). **UNIVERSITÉ DE NAMUR** [BE/BE]; Rue de Bruxelles 61, Namur, 5000 (BE). **UNIVERSITÉ CATHOLIQUE DE LOUVAIN** [BE/BE]; Place de l'Université 1, 1348 Louvain-la-Neuve (BE).

(72) Inventors: **VAN DONINCK, Karine**; Rue du Houblon, 63, 1000 Bruxelles (BE). **NICOLAS, Emilien**; Rue Tumet, 13, 5530 Dorinne (BE). **HALLET, Bernard**; Rue Humerec, 50, 5140 Sombreffe (BE).

(74) Agent: **DE CLERCQ & PARTNERS**; Edgard Gevaertdreef 10a, 9830 Sint-Martens-Latem (BE).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:
— *of inventorship (Rule 4.17(iv))*

Published:
— *with international search report (Art. 21(3))*
— *with sequence listing part of description (Rule 5.2(a))*

(54) Title: METHODS AND COMPOSITIONS FOR INCREASING STRESS TOLERANCE OF CELLS AND ORGANISMS

(57) Abstract: The present application provides an in vitro method for increasing tolerance of a cell to a stress factor, such as radiation, wherein said method comprises genetically engineering the cell to express a specific DNA ligase or a biologically active fragment thereof. Also genetically engineered cells that express the specific DNA ligase or a biologically active fragment thereof are provided, as well as recombinant polynucleotides comprising a polynucleotide that encodes the specific DNA ligase or a biologically active fragment thereof. Further, the genetically engineered cells are provided for use as a medicament or in a pharmaceutical composition. Finally, also method of using the genetically engineered cells in an environment characterized by radiation that is higher than terrestrial background is provided.

WO 2023/161266 A1

METHODS AND COMPOSITIONS FOR INCREASING STRESS TOLERANCE OF CELLS AND ORGANISMS

FIELD OF THE INVENTION

The present invention concerns methods and compositions for increasing tolerance of a cell or an organism to a stress factor, such as radiation. The present invention also provides cells with an increased tolerance to a stress factor and *inter alia* the therapeutic use of such cells is provided. The invention is broadly applicable in different fields, such as the medical field and the space exploration field.

BACKGROUND OF THE INVENTION

All living organisms endure constant assaults arising from endogenous and exogenous damaging conditions. Under a stressful environment, the survival potential of an individual depends on the capacity of protecting its cellular components from the aggression and repairing the damages encountered, the latter one depends on a global signaling network referred to as the DNA damage response (DDR) (Ciccia and Elledge, Mol Cell, 40: 179-204, 2010). DDR senses different types of DNA damages and coordinates a response that includes cell cycle control, activation of transcription, DNA repair pathways, senescence, and cell death. The outcome of the DDR depends on several factors including the number and the complexity of the DNA damages.

Faced with similar DNA damaging conditions, survival capacity varies greatly among living organisms, some being highly sensitive and others being highly tolerant. This continuum reflects the evolution of mechanisms involved in DNA protection and repair that are more efficient, or even novel, in certain organisms compared to others.

Organisms that are tolerant to extremely harsh environments are often found within the archaea and bacteria, such as for example the bacterial species *Deinococcus radiodurans* (Daly, DNA Repair 11, 12-21, 2012) tolerating complete desiccation and high doses of ionizing radiation. However, some animals also exhibit a high DNA damage tolerance when experiencing desiccation or exposure to high doses of ionizing radiation (IR). The most well-known animals in this regard include tardigrades and bdelloid rotifers (Horikawa et al., Int J Radiation Biol, 82: 843-848, 2006; Gadyshey and Meselson, Proc Natl Acad Sci, 105: 5139-5144, 2008). Desiccation and IR create a strong oxidizing environment producing a wide range of DNA damages from multiple base lesions to single-strand breaks (SSBs), and even harmful double-strand breaks (DSBs), causing cell death in most animals.

Tardigrades have led to considerable interest because of their ability to resist multiple stresses, with molecular analyses being advanced by the sequencing of the genomes of several species. Studies

performed on one of the most stress-tolerant tardigrade species *Ramazzottius varieornatus,* highlighted a unique DNA protecting protein (Dsup) (Hashimoto et al. Nat Commun, 7: 12808-14, 2016; Hashimoto et al. Life 7: 26-11, 2017). This highly charged protein binds to chromatin and decreases the DNA fragmentation induced by X-ray irradiation or by reactive oxygen species (ROS).

Bdelloid rotifers are microscopic animals living in semi-terrestrial habitats where they experience frequent desiccation at any stage of their life cycle. This desiccation tolerance predicates their tolerance to ionizing radiation, which is particularly striking as a bdelloid rotifer like *Adineta vaga* can withstand doses up to 3 kGy of X-ray, proton or Fe ions, though doses exceeding 0.5-1.0 kGy render them sterile (Hespeels et al., Front Microbiol, 11: 1792, 2020; Ricci et al. Integr Comp Biol, 45: 759-762, 2005). At such high radiation doses, unlike tardigrades, bdelloid rotifers experience considerable DNA damages including DNA fragmentation created by DSBs. However, after radiation, the organisms manage to recover DNA integrity by an active DNA repair process as observed through the assembly of high molecular weight DNA fragments (Hespeels et al., J Evol Biol, 27: 1334-1345, 2014; Hespeels et al, Front Microbiol, 11: 1792, 2020). While bdelloid rotifers, like tardigrades, were discovered more than 200 years ago, few molecular analyses have yet been performed to identify the key actors of their extreme stress tolerance.

SUMMARY

As corroborated by the experimental section, the inventors developed methods and compositions to increase the resistance of a cell to a stress factor, such as radiation. In particular, it was found that the presence of a specific DNA ligase in a cell increases the tolerance of said cell to a stress factor, showing an increased survival to radiation for example. More specifically, the inventors found that the AvLigE DNA ligase of *Adineta vaga* (*A. vaga*) is an atypical and horizontally acquired DNA ligase with a structural core similar to prokaryotic DNA ligase E and that it strongly contributes to the ligation activity within *A. vaga.* Further, heterologous expression of said DNA ligase within cells significantly improved their stress response, in particular their survival was increased when exposed to ionizing radiation and the *in vitro* DNA ligation activity of transformed cells was higher than control cells. This finding is surprising since heterologous (over)expression of DNA polymerase beta, a gene that is also involved in the DNA Damage Response (DDR) and DNA damage repair, has been shown to cause detrimental effects in mammalian cells (Chan et al., Mutagenesis, 22 (3): 183-188; Zheng et al., Chin. Sci. Bull ; 2013 ; 58 : 3274-3279; Frosina, Eur J Biochem, 200, 267: 2135-2149).

In a first aspect, the application discloses an *in vitro* method for increasing tolerance of a cell to a stress factor. Said method comprises genetically engineering the cell to express a DNA ligase that comprises an amino acid sequence with at least 70% sequence similarity to SEQ ID NO: 1 or SEQ ID NO: 2, or

genetically engineering the cell to express a biologically active fragment of said DNA ligase, wherein the cell is a non-bdelloid cell. In some embodiments, the stress factor is an exogenous stress factor, causes DNA damage in the cell, causes DNA fragmentation in the non-bdelloid cell, causes generation of reactive oxygen species (ROS) in the cell, is radiation, is ionizing radiation, is desiccation, freezing, or any combination thereof.

In another aspect, the application provides a cell genetically engineered to express a DNA ligase that comprises an amino acid sequence with at least 70% sequence similarity to SEQ ID NO: 1 or SEQ ID NO: 2, or cell genetically engineered to express a biologically active fragment of said DNA ligase; or a cell population comprising said cell, wherein the cell is a non-bdelloid cell.

In still another aspect, a non-human transgenic animal is provided, said animal genetically engineered to express a DNA ligase that comprises an amino acid sequence with at least 70% sequence similarity to SEQ ID NO: 1 or SEQ ID NO: 2, or genetically engineered to express a biologically active fragment of said DNA ligase, in at least some of its cells, tissues or organs.

Further, a recombinant polynucleotide is provided, said recombinant polynucleotide comprising a polynucleotide that encodes a DNA ligase that comprises an amino acid sequence with at least 70% sequence similarity to SEQ ID NO: 1 or SEQ ID NO: 2, or that encodes a biologically active fragment of said DNA ligase, wherein the polynucleotide that encodes the DNA ligase is operably linked to one or more heterologous regulatory sequences configured to facilitate the expression of the DNA ligase or fragment thereof in a non-bdelloid cell or an animal. Also provided is a vector comprising said recombinant polynucleotide.

In some embodiments, the non-bdelloid cell is a non-bdelloid animal cell, preferably a warm-blooded animal cell, more preferably a mammalian cell. In some particularly preferred embodiments, the cell is a human cell.

In some embodiments, the DNA ligase is derived from Bdelloidea or bdelloid rotifers, preferably from Adinetidae, more preferably from *Adineta vaga*, even more preferably the DNA ligase is an *Adineta vaga* Ligase E (AvLigE).

In some embodiments, the cell is an immune cell, preferably a T cell or an NK cell, more preferably a cytotoxic T cell.

In a further aspect, the application provides the cell or cell population as disclosed herein for use as a medicament; in particular for use in a method of treating a proliferative disease, or a pharmaceutical composition comprising a cell or cell population as disclosed herein.

In still another aspect, an *in vitro* method is provided of using a non-bdelloid cell or non-bdelloid cell population or a transgenic animal as disclosed herein in an environment characterized by radiation higher than terrestrial background radiation.

Further provided is a method of using a transgenic animal as disclosed herein in an environment characterized by radiation higher than terrestrial background radiation, such as a method of using the transgenic animal as disclosed herein in outer space, such as in deep space exploration. In a specific embodiment, said method is not for treatment of the human or animal body by surgery or therapy.

These and further aspects and preferred embodiments of the invention are described in detail in the following sections and in the appended claims. The subject-matter of the appended claims is hereby specifically incorporated in this specification.

BRIEF DESCRIPTION OF DRAWINGS

The following description of the figures of specific embodiments of the invention is merely exemplary in nature and is not intended to limit the present teachings, their application or uses.

Figure 1. Upregulation of a DNA ligase upon irradiation of the bdelloid rotifer *A. vaga* at 1 kGy of X-ray. A. Global proteomic analysis. Total proteins were extracted from the rotifers at t0h, t4h, t24h, and t72h post-irradiation. Differential expression analysis of each protein was performed with Peaks studio. This scheme shows the proteins that are either upregulated (≥2.5x) or downregulated (≤2.5x) in comparison to the non-irradiated control (NI). **B.** PFGE analysis showing the kinetics of DSB DNA repair upon irradiation. **C.** Western-blot analysis showing the expression pattern of the DNA ligase copy B upon irradiation.

Figure 2. Upregulated DNA ligase in *A. vaga* derives from DNA ligase E. The sequence of the DNA ligase was compared to the available bdelloid rotifer sequences and to the UniProt Reference Clusters (UniRef) 50 database (https://www.uniprot.org/help/uniref) using blastp (evalue 1e-05) and the best 150 hits were extracted to perform an alignment using the MAFFT (https://mafft.cbrc.jp/alignment/software/) multiple sequence alignment program and a phylogenetic analysis using the IQ-TREE program (http://www.iqtree.org/). Representative sequences of DNA ligases K, 1, 3, 4, B, C, and D were manually added to the analysis. The functional domains of the proteins were identified using Interpro scan (https://www.ebi.ac.uk/interpro/search/sequence/) and the structures were manually drawn for each representative case. Only bootstrap values > 50 are indicated on the tree. Numbers between brackets represent the number of species of the same kingdom that are collapsed at each branch.

Figure 3. Relative expression analysis of DNA ligases upon irradiation. A.-C. Kinetics expression analysis of DNA ligase E from *A. vaga*, *A. ricciae*, and *M. verticillata*. The bdelloid rotifers were irradiated at 1 kGy and *M. verticillata* was irradiated at 0.2 kGy of X-Ray. The mRNA level at t0h, t4h, t24h, t48h, and t72h post-irradiation was determined by qPCR and the relative expression were measured in comparison to the non-irradiated control (NI). *M. verticillata* has only one copy of DNA ligase E. **D.** Relative expression analysis of AvLigE in comparison to the other DNA ligases of *A. vaga*. The rotifers were irradiated at 1 kGy and the mRNA levels were measured at t4h post-irradiation. Indicated fold changes correspond to the level of expression of each mRNA in comparison to the level of the same mRNA in the non-irradiated condition (NI). AvLig3 and AvLig4 have only one copy within the genome of *A. vaga*. The histograms represent the average values and standard deviations of three technical replicates.

Figure 4. *In vitro* DNA ligation assay to assess the functionality of AvLigE protein. A. Graphical scheme representing the workflow of immuno-depletion of AvLigE from *A. vaga* protein extracts. The rotifers were irradiated at 1 kGy to stimulate the production of AvLigE-B and the proteins were extracted at t72h post-irradiation. After extraction, AvLigE-B was immuno-depleted from the extracts. A non-irradiated control was also added to the analysis (NI). **B.** Western-blot analysis demonstrating the efficiency of AvLigE-B immuno-depletion. See text for details. **C.** Graphical scheme describing the experimental workflow. A complex of two oligonucleotides is assembled to form a hairpin loop substrate which is covalently bound by one end to wells of 96-well plates and contains a DNA nick at one position to mimic a single-strand break (SSB). A fluorescein moiety is incorporated at the opposite end of the oligonucleotide structure. The complex is then incubated in the presence of increasing concentrations of the protein extracts, followed by a denaturation of the oligonucleotide structure. In the absence of any ligation activity, the oligonucleotide carrying fluorescein is lost from the wells during the denaturation process. Conversely, in the presence of a ligation activity, the DNA nick is sealed and the fluorescein moiety is kept in the wells after denaturation. Therefore, the enzyme activity is determined by quantifying the amount of fluorescein retained in the wells as a result of the denaturation step. Figure adapted from Healing *et al.* (Nucleic Acid Res 47, e61, 2019). Flc, fluorescein. **D.** DNA ligation activities with increasing concentrations of the different *A. vaga* protein extracts. Data represent the mean ± SD of quadruplicate technical replicates. Differential activities were measured by comparison of the slopes of the relative absorbance at 450 nm in the linear portions of the curves. IR, protein extract from irradiated rotifers; IR (AvLigE-B depl.), protein extract from irradiated rotifers and immuno-depleted for AvLigE-B; NI, protein extract from non-irradiated rotifers; NI (AvLigE-B depl.), protein extract from non-irradiated rotifers and immuno-depleted for AvLigE-B.

Figure 5. AvLigE improves X-Ray tolerance of human cell lines. A. Graphical scheme describing the HEK 293T derivative cell lines expressing the AvLigE variants. The control cell line (HEK SCR) includes the empty pDNR plasmid. The other two cell lines express either *AvLigE-B* gene or *AvLigE-A* gene under the constitutive CMV promoter. All constructed cell lines are resistant to puromycin. **B.** *In vitro* DNA ligation assay with the different variants of the HEK 293T cell line. Differential activities were measured by doing a ratio of the slopes in the linear portions of the activity curves. n = 4. **C.** Survival fraction (logarithmic scale) of the different cell lines upon X-Ray irradiation as assessed by colony formation assay. Cells were irradiated at 0.002 or 0.004 kGy. Survival was expressed in comparison to the fraction of surviving cells in the non-irradiated condition. Statistical analysis was performed by multiple t-tests using the Holm-Sidak method. Significant values: * = p value ≤ 0.05; *** = p value ≤ 0.001; **** = p value ≤ 0.0001. n = 6. NI, non-irradiated control. **D.** Representative microscopy images of the colonies formed by the variable cell lines in the different conditions of irradiation. NI, non-irradiated control.

DESCRIPTION OF EMBODIMENTS

As used herein, the singular forms "a", "an", and "the" include both singular and plural referents unless the context clearly dictates otherwise.

The terms "comprising", "comprises" and "comprised of" as used herein are synonymous with "including", "includes" or "containing", "contains", and are inclusive or open-ended and do not exclude additional, non-recited members, elements or method steps. The terms also encompass "consisting of" and "consisting essentially of", which enjoy well-established meanings in patent terminology.

The recitation of numerical ranges by endpoints includes all numbers and fractions subsumed within the respective ranges, as well as the recited endpoints.

The terms "about" or "approximately" as used herein when referring to a measurable value such as a parameter, an amount, a temporal duration, and the like, are meant to encompass variations of and from the specified value, such as variations of +/-10% or less, preferably +/-5% or less, more preferably +/-1% or less, and still more preferably +/-0.1% or less of and from the specified value, insofar such variations are appropriate to perform in the disclosed invention. It is to be understood that the value to which the modifier "about" refers is itself also specifically, and preferably, disclosed.

Whereas the terms "one or more" or "at least one", such as one or more members or at least one member of a group of members, is clear per se, by means of further exemplification, the term encompasses *inter alia* a reference to any one of said members, or to any two or more of said

members, such as, e.g., any ≥3, ≥4, ≥5, ≥6 or ≥7 etc. of said members, and up to all said members. In another example, "one or more" or "at least one" may refer to 1, 2, 3, 4, 5, 6, 7 or more.

The discussion of the background to the invention herein is included to explain the context of the invention. This is not to be taken as an admission that any of the material referred to was published, known, or part of the common general knowledge in any country as of the priority date of any of the claims.

Throughout this disclosure, various publications, patents and published patent specifications are referenced by an identifying citation. All documents cited in the present specification are hereby incorporated by reference in their entirety. In particular, the teachings or sections of such documents herein specifically referred to are incorporated by reference.

Unless otherwise defined, all terms used in disclosing the invention, including technical and scientific terms, have the meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. By means of further guidance, term definitions are included to better appreciate the teaching of the invention. When specific terms are defined in connection with a particular aspect of the invention or a particular embodiment of the invention, such connotation is meant to apply throughout this specification, i.e., also in the context of other aspects or embodiments of the invention, unless otherwise defined.

In the following passages, different aspects or embodiments of the invention are defined in more detail. Each aspect or embodiment so defined may be combined with any other aspect(s) or embodiment(s) unless clearly indicated to the contrary. In particular, any feature indicated as being preferred or advantageous may be combined with any other feature or features indicated as being preferred or advantageous.

Reference throughout this specification to "one embodiment", "an embodiment" means that a particular feature, structure or characteristic described in connection with the embodiment is included in at least one embodiment of the present invention. Thus, appearances of the phrases "in one embodiment" or "in an embodiment" in various places throughout this specification are not necessarily all referring to the same embodiment, but may. Furthermore, the particular features, structures or characteristics may be combined in any suitable manner, as would be apparent to a person skilled in the art from this disclosure, in one or more embodiments. Furthermore, while some embodiments described herein include some but not other features included in other embodiments, combinations of features of different embodiments are meant to be within the scope of the invention, and form different embodiments, as would be understood by those in the art. For example, in the appended claims, any of the claimed embodiments can be used in any combination.

The present invention is at least based on the discovery that the resistance and tolerance of a cell to a stress factor, such as radiation, can be improved by the expression of a specific DNA ligase in said cell. As corroborated by the experimental section, the inventors found that the presence of a specific DNA ligase in a cell increases the tolerance of said cell showing an increased survival to radiation for example, to stress inducing DNA damages. More specifically, the inventors found a specific DNA ligase to be strongly up-regulated in the bdelloid rotifer *A. vaga* upon radiation. *A. vaga* are organisms that manage to recover DNA integrity by an active DNA repair process after exposure to high doses of radiation and the increased expression of the DNA ligase suggests a potential critical role in its DNA repair process. Indeed, the inventors further found that the AvLigE DNA ligase of *A. vaga* is an atypical and horizontally acquired DNA ligase with a structural core similar to prokaryotic DNA ligase E and that it has a strong ligation activity. Furthermore, heterologous expression of said DNA ligase within non-bdelloid cells significantly improved their stress response, in particular their survival was increased when exposed to ionizing radiation and the *in vitro* DNA ligation activity of transformed cells was higher than control cells. These findings are surprising since heterologous (over)expression of DNA polymerase beta, a gene that is also involved in the DNA Damage Response (DDR) and DNA damage repair, has been shown to cause detrimental effects in mammalian cells (Chan et al., Mutagenesis, 22 (3): 183-188; Zheng et al., Chin. Sci. Bull ; 2013 ; 58 : 3274-3279; Frosina, Eur J Biochem, 200, 267: 2135-2149).

The present invention thus provides in some aspects an *in vitro* method for increasing the tolerance of a cell to a stress factor. The method is characterized by genetically engineering the cell to express a DNA ligase that comprises an amino acid sequence with at least 70% sequence similarity to SEQ ID NO:1 or SEQ ID NO: 2, or by genetically engineering the cell to express a biologically active fragment of said DNA ligase, wherein the cell is a non-bdelloid cell.

As used herein, the "tolerance", or also referred to as the "resistance", of a cell or organism to a stress factor is to be understood as the ability of a cell or organism not to be affected or only slightly to be affected by exposure to a particular stress factor which would otherwise be detrimental to the normal functioning of the cell or organism. The terms "tolerance" and "resistance" are known in the art. The term "stress" refers to any non-optimal condition for growth, development or reproduction of the cell or organism. The term "stress factor" is to be understood herein as any factor that causes stress, thus any factor that causes a non-optimal condition for growth, development or reproduction of the cell or organisms. Stress factors thus include all factors which are detrimental to the normal functioning of the cell or organism. Stress factors can be endogenous or exogenous (also called environmental), wherein an endogenous stress factor is to be understood as a stress factor that originates from the

cell or organism, and an exogenous stress factor is to be understood as a stress factor that originates from outside the cell or organism. Examples of environmental or exogenous stress factors include radiation, mechanical stress, heat stress, stress due to limited oxygen, stress due to oxygen radicals, pH stress and osmotic stress.

In some embodiments of the invention, the stress factor is an exogenous stress factor. Exogenous stress factors include radiation such as ionizing radiation; a temperature that is outside of the normal range, such as freeze temperature or extreme warm temperature; suboptimal oxygen availability; osmotic pressure; desiccation; extremes of pH; or any combination of such conditions. In some further preferred embodiments, the stress factor as taught herein is an exogenous stress factor selected from radiation, such as ionizing radiation, desiccation, freezing, or any combination thereof.

Thus, in some embodiments, the *in vitro* method as disclosed herein is for increasing tolerance of a non-bdelloid cell or organism to an exogenous stress factor. Preferably, the exogenous stress factor is selected from radiation, desiccation, freezing, or any combination thereof. In some preferred embodiments, the exogenous stress factor is ionizing radiation. In some other preferred embodiments, the stress factor is desiccation. In still some other preferred embodiments, the stress factor is freezing.

In some embodiments, the stress factor can be an endogenous stress factor, such as for example endogenously produced reactive oxygen species (ROS).

In some embodiments, the stress factor, preferably an exogenous stress factor, causes DNA damage in the cell. The term "DNA damage" as used herein is to be understood as any change or abnormality in the chemical structure of DNA. DNA damage causes changes in the structure of the DNA and thereby prevents the replication mechanism from functioning and performing properly and triggers the DNA damage response (DDR). DNA damage is thus to be understood as an alteration in the chemical structure of DNA, such as base damages (including base alkylation or base loss), nucleotide damages, mismatch, multiple base lesions, DNA crosslinking (including cyclobutane pyrimidine dimers and 6-4 photoproducts), bulky adducts, 8-oxo-7,8-dihydroguanine (8-oxoG), single-strand breaks (SSBs), or double-strand breaks (DSBs). DNA damages can occur via naturally/endogenous factors or via environmental/exogenous factors. In some preferred embodiments of the invention, the DNA damage is caused by an exogenous stress factor, such as radiation, desiccation, freezing, or a combination thereof. The DDR senses different types of DNA damages and coordinates a response that includes cell cycle control, activation of transcription, DNA repair pathways, senescence, and cell death. The outcome of the DDR depends on several factors including the number and the complexity of the DNA damages.

In some embodiments, the stress factor, preferably an exogenous stress factor, causes DNA fragmentation in the cell. The term "DNA fragmentation" as used herein is to be understood as the separation or breaking of DNA strands into pieces. Spontaneous or accidental DNA fragmentation is fragmentation that gradually accumulates in a cell. DNA fragmentation can also be induced by an exogenous factor, such as radiation.

In some embodiments, the stress factor, preferably an exogenous stress factor, causes generation of reactive oxygen species (ROS) in the cell. ROS are to be understood as a group of short-lived, highly reactive, oxygen-containing molecules that can induce DNA damages and trigger the DDR mechanism. Thus, in some embodiments, the stress factor causes generation of ROS in the cell, which on their turn result in DNA damages and/or DNA fragmentation in the cell. In some further embodiments, the exogenous stress factor that causes generation of ROS in the cell is radiation, such as ionizing radiation.

As used herein, the term "ionizing radiation" refers to radiation that consists of subatomic particles or electromagnetic waves that have sufficient energy to ionize atoms or molecules by detaching electrons from them. The particles generally travel at a speed that is greater than 1% of that of light, and the electromagnetic waves are on the high-energy portion of the electromagnetic spectrum. Ionizing radiation can be selected from gamma rays, X-rays, and the higher energy ultraviolet part of the electromagnetic spectrum. Typical ionizing subatomic particles include alpha particles, beta particles, and neutrons. These are typically created by radioactive decay and almost all are energetic enough to ionize. Cosmic rays are an abundant source of ionizing radiation in space, and may also produce radioisotopes on earth, which in turn decay and emit ionizing radiation. Cosmic rays and the decay of radioactive isotopes are the primary sources of natural ionizing radiation on earth. On the other hand, ionizing radiation can also be generated artificially by X-ray tubes, particle accelerators, and nuclear fission.

As disclosed herein, genetic engineering of the cell, in particular the non-bdelloid cell, thus results in an increase of the tolerance of the cell to a stress factor. In some embodiments, the tolerance of the cell to the stress factor is increased at least 1.05-fold, preferably at least 1.10-fold, such as at least 1.5-fold, at least 1.6-fold, at least 1.7-fold, at least 1.8-fold, at least 1.9-fold, at least 2.0-fold, or at least 2.5-fold, compared to a non-genetically engineered cell or to a mock-genetically engineered cell, as measured by cell survival fraction. In other words, the tolerance of the cell to the stress factor is increased by at least 5%, preferably by at least 10%, such as by at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 100%, or even by at least 150%, as compared to a non-genetically engineered cell or to a mock-genetically engineered cell, as measured by cell survival fraction. The cell survival fraction as used herein is to be understood as the

number of viable cells that are present after exposure to the stress factor relative to the number of viable cells that were present before exposure to the stress factor. A non-genetically engineered cell is to be understood as a cell that was not genetically engineered. A mock-genetically engineered cell is to be understood as a cell that underwent all the steps of the genetic engineering similar to the genetically engineered cell to express the DNA ligase or a biologically active fragment thereof as taught herein, except for introduction of the sequence of the DNA ligase or the biologically active fragment thereof. For example, genetic engineering can be performed with an empty vector.

Aspects of the invention thus provide *in vitro* methods characterized by genetically engineering of a cell, more specific a non-bdelloid cell, to express a DNA ligase that comprises an amino acid sequence with at least 70% sequence similarity to SEQ ID NO: 1 or SEQ ID NO: 2, or by genetically engineering the cell to express a biologically active fragment of said DNA ligase. Also provided is a non-bdelloid cell that is genetically engineered to express a DNA ligase that comprises an amino acid sequence with at least 70% sequence similarity to SEQ ID NO: 1 or SEQ ID NO: 2, or a non-bdelloid cell that is genetically engineered to express a biologically active fragment of said DNA ligase. In another aspect, a non-human transgenic animal is provided, said animal genetically engineered to express a DNA ligase that comprises an amino acid sequence with at least 70% sequence similarity to SEQ ID NO: 1 or SEQ ID NO: 2, or said animal genetically engineered to express a biologically active fragment of said DNA ligase, in at least some of its cells, tissues or organs. Further provided is a recombinant polynucleotide comprising a polynucleotide that encodes a DNA ligase that comprises an amino acid sequence with at least 70% sequence similarity to SEQ ID NO: 1 or SEQ ID NO: 2, or that encodes a biologically active fragment of said DNA ligase, operably linked to one or more heterologous regulatory sequences configured to facilitate the expression of the DNA ligase or fragment thereof in a cell or an animal; or a vector comprising said recombinant polynucleotide.

In some embodiments, the DNA ligase as taught herein comprises, consists essentially of or consists of an amino acid sequence with at least 70% sequence similarity to SEQ ID NO: 1:

MSTAATQVLLAETWSEDIDPTGWWMSEKLDGVRAYWSGSNFYSRQGNLFHVPDFFKAALPKVPLDGEIWCGRG
LFQKCISIIKKQGNKVVPDDYKLLTYLIFDAPTHGGKYEDRVKWLQTNVPQDDDKCYASVVGIKKCQNLADLKQCLA
TVNDAGGEGIMLRKPGSLYENKRSSTLLKVKTFYDEEALVVGHKPGKGNCTGVLGALQCQLPNGKRFDVGSGFDM
SQRRNPPKKGSVITFKFQELSNAGIPRFPVFLRIRSDLTWNDVLEAAKTKTPISTTQKVVPSAKLSKQHSILFSIIPSRD
VKTGKKVVASDEEGEQISSSSPSTSTSKTKTDNQEVCQYGAKCYRTNADHLKQYQHPTSSKLTKSKSKTKSSPTSTLKI
TRQSSKKSESQDEPTSPTITSNLLVKDKNKGRHFGDDDKVDDDDDDDEPAATPITSNKRKRNVVEKKPVRKRTKQS
(SEQ ID NO: 1).

In further embodiments, the DNA ligase comprises, consists essentially of or consists of an amino acid sequence with at least 80%, preferably at least 90%, even more preferably at least 95%, at least 96%,

at least 97%, at least 98%, or at least 99% sequence similarity to SEQ ID NO: 1. In particular embodiments, the DNA ligase comprises, consists essentially of or consists of an amino acid sequence with 100% sequence similarity to SEQ ID NO: 1.

In some embodiments, the DNA ligase comprises, consists essentially of or consists of an amino acid sequence with at least 70% sequence identity to SEQ ID NO: 1. In further embodiments, the DNA ligase comprises, consists essentially of or consists of an amino acid sequence with at least 80%, preferably at least 90%, even more preferably at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to SEQ ID NO: 1. In particular embodiments, the DNA ligase comprises, consists essentially of or consists of an amino acid sequence with 100% sequence identity to SEQ ID NO: 1. In certain particularly preferred embodiments, the DNA ligase is as set forth in (i.e., identical to) SEQ ID NO: 1, i.e., identical over the full length thereof to SEQ ID N: 1.

In some embodiments, the DNA ligase comprises, consists essentially of or consists of an amino acid sequence with at least 70% sequence similarity to SEQ ID NO: 2:

MCSEAAKVLLAETWSEDIDPTGWWMSEKLDGVRAYWNGKNFYSRQGNLFHAPDFFKAALPKVPLDGEIWCGR
GLFQKCVSIVKKQANKVVPQDYKLLTYLIFDAPTQGGKYEDRVKWLQANIPQDDDNCYATVVGIKKCEGLAQLKQY
LADVNKAGGEGIMLRKPGSLYEHKRSTTLLKVKTFYDEEAHVIGHKPSKSLIGMTGALECELPNGKRFDVGSGLTM
DQRRKPPKIGSVITFKFQELSNSGSPRFPTFLRVRTDLTWNDVLEAAKTKKPVSVIQKVVPSAKLSKQHSILFSVIPSR
DAKTDKKIVTSDDEDEDAPSTSTSTSTKTNDSRPMCKFGAKCYRTNPDHLKQYQHSSSSTKPAKTTEAESTNKTAKT
AEAESTTKPPCTFGAKCYRTSSIHLATYSHPSKNGTKEASEEDILDTRELVEAEEPASPTTTDNLLVKDKNKGLTFDDD
DDESNDQEMVSRSKKDWTNLENKCKEQSKRLAYLEEMFKTTKRSNSTTDNNEESNKRMKTD (SEQ ID NO: 2).

In some further embodiments, the DNA ligase comprises, consists essentially of or consists of an amino acid sequence with at least 80%, preferably at least 90%, even more preferably at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence similarity to SEQ ID NO: 2. In particular embodiments, the DNA ligase comprises, consists essentially of or consists of an amino acid sequence with 100% sequence similarity to SEQ ID NO: 2.

In some embodiments, the DNA ligase comprises, consists essentially of or consists of an amino acid sequence with at least 70% sequence identity to SEQ ID NO: 2. In further embodiments, the DNA ligase comprises, consists essentially of or consists of an amino acid sequence with at least 80%, preferably at least 90%, even more preferably at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to SEQ ID NO: 2. In particular embodiments, the DNA ligase comprises, consists essentially of or consists of an amino acid sequence with 100% sequence identity to SEQ ID NO: 2. In certain particularly preferred embodiments, the DNA ligase is as set forth in (i.e., identical to) SEQ ID NO: 2, i.e., identical over the full length thereof to SEQ ID N: 2.

The terms "sequence similarity" or "sequence identity" as applied to polypeptides and amino acid sequences means that two peptide sequences, when optimally aligned, share at least some, such as in particular the stated, degree of sequence similarity or identity. The term "sequence identity" with regard to amino acid sequences denotes the extent of overall sequence identity (e.g., including the whole or entire amino acid sequences as recited in the comparison) expressed in % between the amino acid sequences read from N-terminus to C-terminus; and with regard to nucleic acid sequences the extent of overall sequence identity (i.e., including the whole or entire nucleic acid sequences as recited in the comparison) expressed in % between the nucleic acid sequences read from 5'-terminus to 3'-terminus (optionally, complementary sequences may be used in the comparison). Sequence identity may be determined using suitable algorithms for performing sequence alignments and determination of sequence identity as known *per se*.

Where an amino acid sequence differs, varies or diverges from a certain other amino acid sequence – for example, where the former amino acid sequence is said to display a certain degree or percentage of sequence similarity or identity to the latter amino acid sequence, or where the former amino acid sequence is said to differ by a certain number of amino acids from the latter amino acid sequence – such sequence variation may be constituted by one or more amino acid additions (e.g., a single amino acid or a stretch of two or more contiguous amino acids may be added at one position of an amino acid sequence or each independently at two or more positions of an amino acid sequence), deletions (e.g., a single amino acid or a stretch of two or more contiguous amino acids may be deleted at one position of an amino acid sequence or each independently at two or more positions of an amino acid sequence), and/or or substitutions (e.g., a single amino acid or a stretch of two or more contiguous amino acids may substitute a single one or a stretch of two or more contiguous amino acids at one position of an amino acid sequence or each independently at two or more positions of an amino acid sequence).

The term "sequence similarity" is thus broader than the term "sequence identity" since similarity allows conservative modifications, such as substitutions of amino acid residues, having similar physicochemical properties over a defined length of a given alignment. The term "conservative modifications" refers to amino acid modifications that do not significantly affect or alter the functional characteristics of the protein. Such conservative modifications include amino acid substitutions, additions and deletions. A "conservative amino acid substitution" is one in which an amino acid residue is substituted by another amino acid residue having a side chain (R group) with similar chemical properties (e.g., charge or hydrophobicity). In general, a conservative amino acid substitution will not substantially change the functional properties of a protein. In case where two or more amino acid sequences differ from each other by conservative substitutions, the percent or degree of

similarity may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well known to those skilled in the art. See, e.g., Pearson (Methods Mol Biol 24: 307-331; 1994), which is herein incorporated by reference. Examples of groups of amino acids that have side chains with similar chemical properties include 1) aliphatic side chains: glycine, alanine, valine, leucine, and isoleucine; 2) aliphatic-hydroxyl side chains: serine and threonine; 3) amide-containing side chains: asparagine and glutamine; 4) aromatic side chains: phenylalanine, tyrosine, and tryptophan; 5) basic side chains: lysine, arginine, and histidine; 6) acidic side chains: aspartate and glutamate, and 7) sulfur-containing side chains: cysteine and methionine. Preferred conservative amino acids substitution groups are: valine-leucine-isoleucine, phenylalanine-tyrosine, lysine-arginine, alanine-valine, glutamate-aspartate, and asparagine-glutamine. Any substitution of one member of the above-mentioned groups by another member of the same group can be deemed a conservative substitution. By contrast, a non-conservative substitution is a substitution of one amino acid for another with dissimilar characteristics.

Exemplary but non-limiting algorithms to determine sequence identity and similarity include those based on the computer program Basic Local Alignment Search Tool (BLAST), originally described by Altschul *et al.* 1990 (J Mol Biol 215: 403-10), such as the "Blast 2 sequences" algorithm described by Tatusova and Madden 1999 (FEMS Microbiol Lett 174: 247-250), especially BLASTP (for polypeptides) or TBLASTN (for nucleotides), using default parameters. For example, using the published default settings or other suitable settings (such as, e.g., for the BLASTN algorithm: cost to open a gap = 5, cost to extend a gap = 2, penalty for a mismatch = -2, reward for a match = 1, gap x_dropoff = 50, expectation value = 10.0, word size = 28; or for the BLASTP algorithm: matrix = Blosum62 (Henikoff et al., 1992, Proc. Natl. Acad. Sci., 89:10915-10919), cost to open a gap = 11, cost to extend a gap = 1, expectation value = 10.0, word size = 3).

An example procedure to determine the percent identity and similarity between a particular amino acid sequence and a query amino acid sequence will entail aligning the two amino acid sequences each read from N-terminus to C-terminus using the BLASTP program and Blast 2 sequences (Bl2seq) algorithm, available as a web application (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&BLAST_SPEC=blast2seq&LINK_LOC=blasttab) or as a standalone executable program (BLAST version 2.11.0+) at the NCBI web site (https://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/), using suitable algorithm parameters. The BLASTP program using the BLOSUM 62 matrix uses as default algorithm parameters: matrix = Blosum62, cost to open a gap = 11, cost to extend a gap = 1, expectation value = 10.0, word size = 3) (Henikoff & Henikoff, 1989, Proc Natl Acad Sci USA 89: 10915). If the two compared sequences share identity, then the output will present those regions of identity as aligned sequences.

If the two compared sequences do not share identity, then the output will not present aligned sequences. Once aligned, the number of matches will be determined by counting the number of positions where an identical amino acid residue is presented in both sequences. The percent identity is determined by dividing the number of matches by the length of the query sequence, followed by multiplying the resulting value by 100. The percent identity value may, but need not, be rounded to the nearest tenth. For example, 78.11, 78.12, 78.13, and 78.14 may be rounded down to 78.1, while 78.15, 78.16, 78.17, 78.18, and 78.19 may be rounded up to 78.2. It is further noted that the detailed view for each segment of alignment as outputted by Bl2seq already conveniently includes the percentage of identities.

The present specification discusses certain molecules which may be peptides, polypeptides or proteins, such as for example, a DNA ligase, in particular a DNA ligase derived from Bdelloidea or bdelloid rotifers, preferably from Adinetidae, even more preferably derived from *Adineta vaga*. In preferred embodiments, the DNA ligase is an *Adineta vaga* Ligase E (AvLigE). Even more specifically, the DNA ligase comprises the amino acid sequence SEQ ID NO: 1 or SEQ ID NO: 2.

The term "protein" generally encompasses macromolecules comprising one or more polypeptide chains. The term "polypeptide" generally encompasses linear polymeric chains of amino acid residues linked by peptide bonds. A "peptide bond", "peptide link" or "amide bond" is a covalent bond formed between two amino acids when the carboxyl group of one amino acid reacts with the amino group of the other amino acid, thereby releasing a molecule of water. Especially when a protein is only composed of a single polypeptide chain, the terms "protein" and "polypeptide" may be used interchangeably to denote such a protein. The terms are not limited to any minimum length of the polypeptide chain. Polypeptide chains consisting essentially of or consisting of 50 or less (≤50) amino acids, such as ≤45, ≤40, ≤35, ≤30, ≤25, ≤20, ≤15, ≤10 or ≤5 amino acids may be commonly denoted as a "peptide". In the context of proteins, polypeptides or peptides, a "sequence" is the order of amino acids in the chain in an amino to carboxyl terminal direction in which residues that neighbor each other in the sequence are contiguous in the primary structure of the protein, polypeptide or peptide. The terms may encompass naturally, recombinantly, semi-synthetically or synthetically produced proteins, polypeptides or peptides. Hence, for example, a protein, polypeptide or peptide can be present in or isolated from nature, e.g., produced or expressed natively or endogenously by a cell or tissue and optionally isolated therefrom; or a protein, polypeptide or peptide can be recombinant, i.e., produced by recombinant DNA technology, and/or can be, partly or entirely, chemically or biochemically synthesized. Without limitation, a protein, polypeptide or peptide can be produced recombinantly by a suitable host or host cell expression system and optionally isolated therefrom (e.g.,

a suitable bacterial, yeast, fungal, plant or animal host or host cell expression system), or produced recombinantly by cell-free translation or cell-free transcription and translation, or non-biological peptide, polypeptide or protein synthesis. The terms also encompasses proteins, polypeptides or peptides that carry one or more co- or post-expression-type modifications of the polypeptide chain(s), such as, without limitation, glycosylation, lipidation, acetylation, amidation, phosphorylation, sulphonation, methylation, pegylation (covalent attachment of polyethylene glycol typically to the N-terminus or to the side-chain of one or more Lys residues), ubiquitination, sumoylation, cysteinylation, glutathionylation, oxidation of methionine to methionine sulphoxide or methionine sulphone, signal peptide removal, N-terminal Met removal, conversion of pro-enzymes or pre-hormones into active forms, etc. Such co- or post-expression-type modifications may be introduced *in vivo* by a host cell expressing the proteins, polypeptides or peptides (co- or post-translational protein modification machinery may be native to the host cell and/or the host cell may be genetically engineered to comprise one or more (additional) co- or post-translational protein modification functionalities), or may be introduced *in vitro* by chemical (e.g., pegylation) and/or biochemical (e.g., enzymatic) modification of the isolated proteins, polypeptides or peptides.

The term "amino acid" encompasses naturally occurring amino acids, naturally encoded amino acids, non-naturally encoded amino acids, non-naturally occurring amino acids, amino acid analogues and amino acid mimetics that function in a manner similar to the naturally occurring amino acids, all in their D- and L-stereoisomers, provided their structure allows such stereoisomeric forms. Amino acids are referred to herein by either their name, their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. A "naturally encoded amino acid" refers to an amino acid that is one of the 20 common amino acids or pyrrolysine, pyrroline-carboxy-lysine or selenocysteine. The 20 common amino acids are: Alanine (A or Ala), Cysteine (C or Cys), Aspartic acid (D or Asp), Glutamic acid (E or Glu), Phenylalanine (F or Phe), Glycine (G or Gly), Histidine (H or His), Isoleucine (I or Ile), Lysine (K or Lys), Leucine (L or Leu), Methionine (M or Met), Asparagine (N or Asn), Proline (P or Pro), Glutamine (Q or Gln), Arginine (R or Arg), Serine (S or Ser), Threonine (T or Thr), Valine (V or Val), Tryptophan (W or Trp), and Tyrosine (Y or Tyr). A "non-naturally encoded amino acid" refers to an amino acid that is not one of the 20 common amino acids or pyrrolysine, pyrroline-carboxy-lysine or selenocysteine. The term includes without limitation amino acids that occur by a modification (such as a post-translational modification) of a naturally encoded amino acid, but are not themselves naturally incorporated into a growing polypeptide chain by the translation complex, as exemplified without limitation by N-acetylglucosaminyl-L-serine, N-acetylglucosaminyl-L-threonine, and O-phosphotyrosine. Further examples of non-naturally encoded, un-natural or modified amino acids include 2-Aminoadipic acid, 3-Aminoadipic acid, beta-Alanine,

beta-Aminopropionic acid, 2-Aminobutyric acid, 4-Aminobutyric acid, piperidinic acid, 6-Aminocaproic acid, 2-Aminoheptanoic acid, 2-Aminoisobutyric acid, 3-Aminoisobutyric acid, 2-Aminopimelic acid, 2,4 Diaminobutyric acid, Desmosine, 2,2'-Diaminopimelic acid, 2,3-Diaminopropionic acid, N-Ethylglycine, N-Ethylasparagine, homoserine, homocysteine, Hydroxylysine, allo-Hydroxylysine, 3-Hydroxyproline, 4-Hydroxyproline, Isodesmosine, allo-Isoleucine, N-Methylglycine, N-Methylisoleucine, 6-N-Methyllysine, N-Methylvaline, Norvaline, Norleucine, or Ornithine. Also included are amino acid analogues, in which one or more individual atoms have been replaced either with a different atom, an isotope of the same atom, or with a different functional group. Also included are un-natural amino acids and amino acid analogues described in Ellman et al. Methods Enzymol. 1991, vol. 202, 301-36. The incorporation of non-natural amino acids into proteins, polypeptides or peptides may be advantageous in a number of different ways. For example, D-amino acid-containing proteins, polypeptides or peptides exhibit increased stability *in vitro* or *in vivo* compared to L-amino acid-containing counterparts. More specifically, D-amino acid-containing proteins, polypeptides or peptides may be more resistant to endogenous peptidases and proteases, thereby providing improved bioavailability of the molecule and prolonged lifetimes *in vivo*.

In certain embodiments, the cell, the non-human transgenic animal or its cells, tissues or organs, may express a biologically active fragment or portion of the DNA ligase as described herein, or the polynucleotide encodes a biologically active fragment of the DNA ligase as disclosed herein. The terms "fragment" or "portion" of a protein, polypeptide or peptide generally encompass N-terminally and/or C-terminally deleted or truncated forms of the full-length protein, polypeptide or peptide. The term encompasses fragments arising by any mechanism, such as without limitation, by expression of a truncated form of the full-length protein, polypeptide or peptide, or by physical, chemical or enzymatic proteolysis of the full-length protein, polypeptide or peptide *in vivo* or *in vitro*. As used herein, the term "biologically active fragment" refers to any fragment, derivative or analogue of the DNA ligase that possesses an *in vivo* or *in vitro* activity that is characteristic for the DNA ligase itself, i.e. that possesses also DNA ligating capacity. In some embodiments, the biologically active fragment possesses 10%, 20%, 40%, 50%, 60%, 70%, 80%, 90% or even higher of the activity of the DNA ligase itself in any *in vivo* or *in vitro* assay of interest. It will be apparent to a person of skill in the art that a full-length protein or nucleic acid encoding such or a fragment thereof can be used according to the methods described herein. A portion of a protein is preferably a biologically active fragment thereof. Fragments that are biologically active can be identified according to methods known in the art and using an assay that can monitor the activity of the DNA ligase. Assays for determining the activity of an DNA ligase protein are described, e.g. the *in vitro* DNA ligation assay as described in Healing et al. (Nucleic Acids Res. 47; e61; 2019).

In some embodiments, the biologically active fragment of the DNA ligase as disclosed herein comprises at least an ATP dependent DNA ligase domain and a DNA ligase oligonucleotide binding (OB)-like domain.

The skilled person shall appreciate that a DNA ligase molecule as discussed herein will typically require the inclusion of one or more suitable DNA ligase domains, such as an ATP-dependent ligase domain or a DNA ligase OB-like domain. For example, ATP-dependent DNA ligase domains specifically catalyze the joining of single-stranded breaks (nicks) in dsDNA, whereas a DNA ligase OB-like domain engages the minor groove of the DNA duplex upon nick binding and assists in the adenylation reaction. In preferred embodiments, the DNA ligase as disclosed herein comprises both an ATP-dependent DNA ligase domain and a DNA ligase OB-like domain. In case of a variant of the DNA ligase as taught herein, the skilled person still finds the location of the ATP-dependent DNA ligase domain and of the DNA ligase OB-like domain within said variant. For example, the skilled person would understand where these domains are in the fragment that shows at least 70% sequence similarity with SEQ ID NO: 1 or SEQ ID NO: 2, or any fragments thereof.

In certain embodiments, the DNA ligase as taught herein comprises, consists essentially of or consists of an ATP-dependent DNA ligase domain wherein said domain displays at least 70% identity, preferably at least 80% identity, more preferably at least 85% identity, more preferably at least 90% identity, even more preferably at least 95% identity, such as particularly preferably at least 96%, at least 97%, at least 98% or at least 99% or 100% identity to amino acid positions 19-181 of SEQ ID NO: 1.

In certain embodiments, the DNA ligase as taught herein comprises, consists essentially of or consists of an ATP-dependent DNA ligase domain wherein said domain displays at least 70% identity, preferably at least 80% identity, preferably at least 85% identity, more preferably at least 90% identity, even more preferably at least 95% identity, such as particularly preferably at least 96%, at least 97%, at least 98% or at least 99% or 100% identity to amino acid positions 19-181 of SEQ ID NO: 2.

In certain embodiments, the DNA ligase as taught herein comprises, consists essentially of or consists of a DNA ligase OB-like domain wherein said domain displays at least 70%, preferably at least 80% identity, preferably at least 85% identity, more preferably at least 90% identity, even more preferably at least 95% identity, such as particularly preferably at least 96%, at least 97%, at least 98% or at least 99% or 100% identity to amino acid positions 195-260 of SEQ ID NO: 1.

In certain embodiments, the DNA ligase as taught herein comprises, consists essentially of or consists of a DNA ligase OB-like domain wherein said domain displays at least 70% identity, preferably at least

80% identity, preferably at least 85% identity, more preferably at least 90% identity, even more preferably at least 95% identity, such as particularly preferably at least 96%, at least 97%, at least 98% or at least 99% or 100% identity to amino acid positions 200-259 of SEQ ID NO: 2.

In certain embodiments, the DNA ligase comprises, consists essentially of or consists of (i) an ATP-dependent DNA ligase domain and of (ii) a DNA ligase OB-like domain wherein said domains (i) – (ii) each independently display at least 70% identity, preferably at least 80% identity, preferably at least 85% identity, more preferably at least 90% identity, even more preferably at least 95% identity, such as particularly preferably at least 96%, at least 97%, at least 98% or at least 99% or 100% identity to (i) amino acid positions 19-181 and (ii) amino acid position 195-260 of SEQ ID NO: 1.

In certain embodiments, the DNA ligase comprises, consists essentially of or consists of (i) an ATP-dependent DNA ligase domain and of (ii) a DNA ligase OB-like domain wherein said domain (i) – (ii) each independently display at least 70% identity, preferably at least 80% identity, preferably at least 85% identity, more preferably at least 90% identity, even more preferably at least 95% identity, such as particularly preferably at least 96%, at least 97%, at least 98%, or at least 99% or 100% identity to (i) amino acid positions 19-181 and (ii) amino acid positions 200-259 of SEQ ID NO: 2.

In some embodiments, the ATP dependent DNA ligase domain comprises the amino acid sequence SEQ ID NO: 3 or SEQ ID NO: 4. In certain embodiments, the DNA ligase domain is an ATP dependent DNA ligase domain that comprises, consists essentially of or consists of an amino acid sequence at least 70%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or 100% identical to SEQ ID NO: 3. In certain embodiments, the DNA ligase domain is an ATP dependent DNA ligase domain that comprises, consists essentially of or consists of an amino acid sequence at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or 100% identical to SEQ ID NO: 4.

In some embodiments, the ATP dependent DNA ligase domain comprises, consists essentially of or consists of the amino acid sequence SEQ ID NO: 3:

DPTGWWMSEKLDGVRAYWSGSNFYSRQGNLFHVPDFFKAALPKVPLDGEIWCGRGLFQKCISIIKKQGNKVVPD DYKLLTYLIFDAPTHGGKYEDRVKWLQTNVPQDDDKCYASVVGIKKCQNLADLKQCLATVNDAGGEGIMLRKPGS LYENKRSSTLLKVK (SEQ ID NO: 3).

In some embodiments, the ATP dependent DNA ligase domain comprises, consists essentially of or consists of the amino acid sequence SEQ ID NO: 4:

DPTGWWMSEKLDGVRAYWNGKNFYSRQGNLFHAPDFFKAALPKVPLDGEIWCGRGLFQKCVSIVKKQANKVVP QDYKLLTYLIFDAPTQGGKYEDRVKWLQANIPQDDDNCYATVVGIKKCEGLAQLKQYLADVNKAGGEGIMLRKPG SLYEHKRSTTLLKVK (SEQ ID NO: 4).

In some embodiments, the DNA ligase OB-like domain comprises the amino acid sequence SEQ ID NO: 5 or SEQ ID NO: 6. In certain embodiments, the DNA ligase domain is a DNA ligase OB-like domain that comprises, consists essentially of or consists of an amino acid sequence at least 70%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or 100% identical to SEQ ID NO: 5. In certain embodiments, the DNA ligase domain is a DNA ligase OB-like domain that comprises, consists essentially of or consists of an amino acid sequence at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or 100% identical to SEQ ID NO: 6.

In some embodiments, the DNA ligase OB-like domain comprises, consists essentially of or consists of the amino acid sequence SEQ ID NO: 5:

PGKGNCTGVLGALQCQLPNGKRFDVGSGFDMSQRRNPPKKGSVITFKFQELSNAGIPRFPVFLRIR (SEQ ID NO: 5).

In some embodiments, the DNA ligase OB-like domain comprises, consists essentially of or consists of the amino acid sequence SEQ ID NO: 6:

IGMTGALECELPNGKRFDVGSGLTMDQRRKPPKIGSVITFKFQELSNSGSPRFPTFLRVR (SEQ ID NO: 6).

Without limitation, the biologically active fragment of a protein, polypeptide or peptide may represent at least about 20% (by amino acid number), e.g., at least about 30%, or at least about 40%, of the contiguous amino acid sequence of the full-length protein, polypeptide or peptide. For example, a fragment or portion of one of the illustrative 460 amino acid-long AvLigE-A as set forth in SEQ ID NO: 1 may include a sequence of at least about 150, or at least about 180, or at least about 200, or at least about 300, preferably at least 400 or at least 450 amino acids of the full-length protein as set forth in SEQ ID NO: 1. Or, a fragment or portion of the 520 amino acid-long AvLigE-B as set forth in SEQ ID NO: 2 may include a sequence of at least about 150, or at least about 180, or at least about 200, or at least about 300, preferably at least 400 or at least 450 amino acids of the full-length protein as set forth in SEQ ID NO: 2.

As already discussed, the present invention particularly envisages genetic engineering of cells, in particular non-bdelloid cells, to express a DNA ligase that comprises an amino acid sequence as disclosed herein or to express a biologically active fragment of said DNA ligase. Aspects also relate to

a recombinant polynucleotide comprising a polynucleotide that encodes a DNA ligase as disclosed herein or that encodes a biologically active fragment of said DNA ligase, wherein said polynucleotide is operably linked to one or more heterologous regulatory sequences configured to facilitate the expression of the DNA ligase or fragment thereof in a cell or an animal. Also a vector comprising said recombinant polynucleotide is disclosed herein. By "encoding" is meant that a nucleic acid sequence or part(s) thereof corresponds, by virtue of the genetic code of an organism in question to a particular amino acid sequence, e.g., the amino acid sequence of one or more desired proteins or polypeptides, or to another nucleic acid sequence in a template-transcription product (e.g., RNA or RNA analogue) relationship.

The term "nucleic acid" as used herein typically refers to a polymer (preferably a linear polymer) of any length composed essentially of nucleoside units. A nucleoside unit commonly includes a heterocyclic base and a sugar group. Heterocyclic bases may include *inter alia* purine and pyrimidine bases such as adenine (A), guanine (G), cytosine (C), thymine (T) and uracil (U), which are widespread in naturally-occurring nucleic acids, other naturally-occurring bases (e.g., xanthine, inosine, hypoxanthine), as well as chemically or biochemically modified (e.g., methylated), non-natural or derivatised bases. Exemplary modified nucleobases include, without limitation, 5-substituted pyrimidines, 6-azapyrimidines and N-2, N-6 and O-6 substituted purines, including 2-aminopropyladenine, 5-propynyluracil and 5-propynylcytosine. In particular, 5-methylcytosine substitutions have been shown to increase nucleic acid duplex stability. Sugar groups may include *inter alia* pentose (pentofuranose) groups such as preferably ribose and/or 2-deoxyribose common in naturally-occurring nucleic acids, or arabinose, 2-deoxyarabinose, threose or hexose sugar groups, as well as modified or substituted sugar groups (such as, without limitation, 2'-O-alkylated, e.g., 2'-O-methylated or 2'-O-ethylated sugars such as ribose; 2'-O-alkyloxyalkylated, e.g., 2'-O-methoxyethylated sugars such as ribose; or 2'-O,4'-C-alkylene-linked, e.g., 2'-O,4'-C-methylene-linked or 2'-O,4'-C-ethylene-linked sugars such as ribose; 2'-fluoro-arabinose, etc.). Nucleoside units may be linked to one another by any one of numerous known inter-nucleoside linkages, including *inter alia* phosphodiester linkages common in naturally-occurring nucleic acids, and further modified phosphate- or phosphonate-based linkages such as phosphorothioate, alkyl phosphorothioate such as methyl phosphorothioate, phosphorodithioate, alkylphosphonate such as methylphosphonate, alkylphosphonothioate, phosphotriester such as alkylphosphotriester, phosphoramidate, phosphoropiperazidate, phosphoromorpholidate, bridged phosphoramidate, bridged methylene phosphonate, bridged phosphorothioate; and further siloxane, carbonate, sulfamate, carboalkoxy, acetamidate, carbamate such as 3'-N-carbamate, morpholino, borano, thioether, 3'-thioacetal, and sulfone internucleoside linkages. Preferably, inter-nucleoside linkages may be phosphate-based

linkages including modified phosphate-based linkages, such as more preferably phosphodiester, phosphorothioate or phosphorodithioate linkages or combinations thereof. The term "nucleic acid" also encompasses any other nucleobase containing polymers such as nucleic acid mimetics, including, without limitation, peptide nucleic acids (PNA), peptide nucleic acids with phosphate groups (PHONA), locked nucleic acids (LNA), morpholino phosphorodiamidate-backbone nucleic acids (PMO), cyclohexene nucleic acids (CeNA), tricyclo-DNA (tcDNA), and nucleic acids having backbone sections with alkyl linkers or amino linkers (see, e.g., Kurreck 2003 (Eur J Biochem 270: 1628–1644)). "Alkyl" as used herein particularly encompasses lower hydrocarbon moieties, e.g., C_1-C_4 linear or branched, saturated or unsaturated hydrocarbon, such as methyl, ethyl, ethenyl, propyl, 1-propenyl, 2-propenyl, and isopropyl.

Nucleic acids as intended herein may include naturally occurring nucleosides, modified nucleosides or mixtures thereof. A modified nucleoside may include a modified heterocyclic base, a modified sugar moiety, a modified inter-nucleoside linkage or a combination thereof. The term "nucleic acid" further preferably encompasses DNA, RNA and DNA/RNA hybrid molecules, specifically including hnRNA, pre-mRNA, mRNA, cDNA, genomic DNA, amplification products, oligonucleotides, and synthetic (e.g., chemically synthesized) DNA, RNA or DNA/RNA hybrids. A nucleic acid can be naturally occurring, e.g., present in or isolated from nature, can be recombinant, i.e., produced by recombinant DNA technology, and/or can be, partly or entirely, chemically or biochemically synthesized. A "nucleic acid" can be double-stranded, partly double stranded, or single-stranded. Where single-stranded, the nucleic acid can be the sense strand or the antisense strand. In addition, nucleic acid can be circular or linear.

In certain embodiments, the DNA ligase, such as in particular the DNA ligase as set forth in SEQ ID NO: 1 or SEQ ID NO: 2, may be encoded by a nucleic acid comprising, consisting essentially of or consisting of the nucleic acid sequence set forth in SEQ ID NO: 7 or SEQ ID NO: 8 respectively:

ATGAGTACAGCAGCAACACAGGTTCTTCTTGCAGAAACATGGAGCGAAGATATCGATCCTACTGGTTGGTGG
ATGAGTGAAAAACTTGATGGTGTTCGAGCGTATTGGAGTGGGAGTAACTTTTATTCTCGACAAGGAAATCTTT
TTCACGTGCCTGATTTTTTCAAGGCCGCACTTCCCAAAGTACCCTTGGATGGAGAAATATGGTGTGGTCGTGG
ACTCTTTCAAAAATGTATAAGTATTATAAAGAAACAGGGAAATAAAGTTGTTCCCGATGATTATAAGCTCTTGA
CTTATTTGATTTTCGATGCACCAACTCATGGAGGAAAATATGAAGATCGTGTCAAATGGTTACAGACGAATGT
ACCACAAGATGACGATAAATGTTATGCATCAGTTGTCGGTATAAAAAAATGTCAAAATCTTGCAGATCTTAAA
CAATGTTTAGCTACTGTAAATGATGCCGGTGGAGAAGGAATTATGCTCAGGAAACCTGGTAGTCTATATGAAA
ATAAGCGTTCATCAACATTACTGAAAGTAAAAACGTTTTATGACGAAGAAGCTCTAGTTGTTGGTCATAAACCT
GGAAAAGGAAATTGTACAGGAGTTTTAGGTGCTCTTCAATGTCAATTACCTAATGGAAAACGTTTCGATGTTG
GAAGTGGCTTTGATATGTCTCAACGTCGAAATCCACCTAAAAAAGGTTCAGTGATAACATTTAAATTTCAAGA

ATTATCTAATGCTGGTATTCCACGTTTTCCTGTATTTCTTCGCATTCGTAGTGATTTAACTTGGAATGATGTTCTT
GAAGCAGCTAAAACAAAAACACCAATAAGTACAACACAAAAAGTTGTACCATCCGCAAAATTATCTAAGCAAC
ATTCAATATTATTCTCTATTATTCCATCACGTGATGTTAAAACTGGTAAGAAAGTTGTTGCATCCGATGAAGAA
GGTGAACAAATATCTTCGTCATCACCTTCGACATCAACTTCAAAAACGAAGACAGATAACCAAGAAGTATGTC
AATATGGAGCGAAATGTTATCGAACTAATGCCGATCACCTCAAACAATATCAACATCCAACATCATCTAAATTA
ACGAAATCGAAATCGAAAACAAAATCATCTCCTACCTCTACATTAAAAATTACACGACAAAGTTCTAAGAAATC
AGAATCTCAAGATGAACCAACTTCACCAACTATCACAAGTAATCTACTTGTCAAAGATAAAAATAAAGGTCGTC
ATTTCGGTGACGATGATAAGGTAGATGATGATGATGATGATGAACCAGCTGCAACACCTATAACAAGTA
ATAAACGAAAACGAAATGTTGTTGAAAAAAAACCTGTTAGAAAACGTACAAAGCAATCTTAG (SEQ ID NO: 7);

ATGTGTAGTGAAGCTGCTAAAGTTCTTCTTGCTGAAACATGGTCAGAAGATATTGATCCGACAGGATGGTGGA
TGAGTGAAAAACTCGATGGTGTTCGAGCATACTGGAATGGAAAAAATTTTTATTCAAGACAAGGAAATCTTTT
TCATGCACCTGATTTTTTCAAGGCTGCACTACCTAAAGTTCCACTCGATGGTGAAATATGGTGTGGCCGGGGTT
TATTTCAAAAATGTGTTAGTATTGTCAAAAAACAAGCAAATAAAGTGGTGCCTCAAGATTATAAACTACTGAC
GTATCTAATTTTTGACGCACCAACTCAAGGAGGAAAATATGAAGATCGTGTCAAATGGTTACAAGCAAATATT
CCACAAGATGATGATAACTGTTATGCTACAGTTGTTGGTATTAAAAAATGCGAAGGTTTAGCACAATTAAAAC
AATATTTAGCTGATGTGAACAAAGCTGGTGGAGAAGGTATTATGCTTCGTAAACCAGGTAGTCTTTATGAACA
TAAACGTTCAACTACTTTACTGAAAGTAAAAACGTTTTATGACGAAGAAGCTCATGTCATTGGTCATAAACCGA
GTAAAAGCTTAATTGGTATGACTGGTGCTCTTGAATGTGAATTACCGAATGGAAAACGATTTGATGTTGGTAG
TGGACTTACTATGGATCAACGACGAAAACCACCAAAAATCGGTTCAGTCATTACATTTAAATTTCAAGAATTAT
CCAATAGTGGTAGCCCACGATTTCCGACATTTCTTCGTGTTCGTACTGATCTGACATGGAATGATGTTCTTGAA
GCAGCTAAAACAAAAAAACCTGTCAGTGTCATACAGAAAGTGGTTCCATCGGCAAAATTATCTAAACAACATT
CAATATTATTTTCTGTTATTCCATCACGTGATGCCAAAACCGATAAGAAAATCGTCACATCCGATGACGAAGAT
GAAGATGCACCATCTACATCTACATCCACATCGACAAAGACGAATGATTCTCGACCGATGTGTAAATTTGGTG
CAAAATGTTATCGAACAAATCCTGATCATCTAAAACAATATCAACATTCGTCATCCAGTACGAAACCTGCCAAA
ACTACGGAAGCTGAATCGACCAATAAAACTGCTAAAACTGCAGAAGCAGAATCTACAACCAAACCGCCCTGCA
CATTTGGTGCGAAATGTTATCGTACAAGTTCCATTCATTTAGCTACTTATTCTCATCCATCTAAAAATGGAACAA
AGGAAGCGAGTGAAGAAGATATACTTGATACTCGTGAATTAGTTGAAGCTGAAGAACCTGCTTCACCAACCA
CAACTGATAATTTATTAGTCAAAGATAAAAATAAAGGCCTGACATTCGACGATGATGATGATGAAAGTAATGA
TCAAGAAATGGTTAGTCGTTCTAAAAAAGATTGGACCAATCTTGAAAATAAATGCAAAGAACAAAGCAAACGT
TTAGCATATCTTGAAGAAATGTTTAAAACGACAAAACGATCGAATTCAACAACTGATAATAATGAAGAAAGCA
ACAAACGTATGAAAACAGATTGA (SEQ ID NO: 8).

Optionally, in said nucleic acid sequences any codon may each independently be replaced by another codon encoding the same amino acid, preferably wherein the nucleic acid sequence is thereby codon

optimized for expression in cells of a desired species, such as in human cells, according to codon optimization principles known in the art. For example, wherein the DNA ligase is encoded by a nucleic acid comprising, consisting essentially of or consisting of a nucleic acid sequence at least 60%, at least 70%, at least 80%, at least 90%, or at least 95% identical to SEQ ID NO: 7 or SEQ ID NO: 8. In particular embodiments, the DNA ligase is encoded by a nucleic acid sequence comprising, consisting essentially of or consisting of a nucleic acid sequence that is 100% identical to SEQ ID NO: 7. In particular embodiments, the DNA ligase is encoded by a nucleic acid sequence comprising, consisting essentially of or consisting of a nucleic acid sequence that is 100% identical to SEQ ID NO: 8.

The terms "identity", "sequence identity", or "identical" in the context of nucleotide sequences may be used interchangeably herein, and refer to the extent that nucleic acid sequences are identical on a nucleotide-by-nucleotide basis, over a window of comparison. The percentage of sequence identity may be calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison, and multiplying the result by 100 to yield the percentage of sequence identity. The percent identity value may, but need not, be rounded to the nearest tenth. For example, 98.11, 98.12, 98.13 and 98.14 may be rounded down to 98.1, while 98.15, 98.16, 98.17, 98.18, and 98.19 may be rounded up to 98.2.

Sequence identity between nucleic acids as envisaged herein may be determined using suitable algorithms for performing sequence alignments and determination of sequence identity as known *per se*. Exemplary, but non-limiting algorithms include those based on the Basic Local Alignment Search Tool (BLAST) originally described by Altschul et al. 1990 (J Mol Biol 215: 403-410), such as the "Blast 2 sequences" tool described by Tatusova and Madden 1999 (FEMS Microbiol Lett 174: 247-250), or the "blastn suite-2sequences" sequence alignment algorithm described by Zheng Zhang et al. 2000 (J Comput Biol 2000, 7(1-2): 203-214), now incorporated into the BLAST program suite available at ncbi.nlm.nigh.gov. The skilled person can implement such algorithms and set the requisite parameters. By means of an example and without limitation, parameters for the BLASTN program may be as follows: cost to open a gap = 0, cost to extend a gap = 2.5, reward for a match = 1, penalty for a mismatch = -2, Expect value = 0.05, word size = 25, Low Complexity Filter = Yes.

There are further algorithms known in the art that can be used to measure nucleotide sequence identity. Nucleotide sequence identity can be measured by a local or global alignment, preferably implementing an optimal local or optimal global alignment algorithm. For example, a global alignment

may be generated using an implementation of the Needleman-Wunch algorithm (Needleman & Wunsch, 1970, J Mol Biol 48(3): 443-453). For example, a local alignment (which does not consider the entirety of the sequence but tries to find the longest subsequence that confirms to a given matching criteria) may be generated using an implementation of the Smith-Waterman algorithm (Smith & Waterman, 1981, J Mol Biol 147(1): 195-197). Optimal global alignments using the Needleman-Wunsch algorithm and optimal local alignments using the Smith-Waterman algorithm are implemented in USEARCH (https://drive5.com/usearch/), for example USEARCH version 11.0.667.

A gap is a region of an alignment wherein a sequence does not align to a position in the other sequence of the alignment. In global alignments, terminal gaps are discarded before identity is calculated. For both local and global alignments, internal gaps are counted as differences. A terminal gap is a region beginning at the end of a sequence in an alignment wherein the nucleotide in the terminal position of that sequence does not correspond to a nucleotide position in the other sequence of the alignment and extending for all contiguous positions in that sequence wherein the nucleotides of the sequence do not correspond to a nucleotide position in the other sequence of the alignment.

Nucleic acid sequence identity as envisaged herein in particular denotes overall nucleic acid sequence identity, i.e., sequence identity calculated from optimally aligning the whole sequences of the to-be-compared DNA ligase genes. In other words, the nucleic acid sequence to be aligned is the DNA ligase gene, and the window of comparison corresponds to the whole region of optimal alignment between the complete DNA ligase sequence, i.e. , to the alignment length. Hence, in an example, a query DNA ligase gene sequence, such as the sequence set forth in SEQ ID NO: 7, is optimally aligned with another DNA ligase gene sequence (in case of a global alignment any terminal gaps are disregarded; in case of a local alignment the region of alignment will be expressed as the region between a given 5' and a given 3' position in the query sequence), and the percentage sequence identity is calculated over a window of comparison which corresponds to the whole region of alignment, counting internal gaps as differences.

The polynucleotide as envisaged herein preferably comprises a DNA ligase polynucleotide sequence of which the length is between 90% and 110%, more preferably between 95% and 105% of the length of the DNA ligase polynucleotide shown in SEQ ID NO: 7 or SEQ ID NO: 8. These sequences can be subjected to pairwise sequence comparisons with SEQ ID NO: 7 or SEQ ID NO: 8.

The term "substantial identity" or "substantially identical", when referring to a nucleic acid or fragment thereof, indicates that, when optimally aligned with appropriate nucleotide insertions or deletions with another nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 90%, and more preferably at least about 95%, 96%, 97%, 98% or 99% of the

nucleotide bases, as measured by any well-known algorithm of sequence identity, such as FASTA, BLAST or GAP. A nucleic acid molecule having substantial identity to a reference nucleic acid molecule may, in certain instances, encode a polypeptide having the same or substantially similar amino acid sequence as the polypeptide encoded by the reference nucleic acid molecule.

To allow for expression of a nucleic acid encoding a DNA ligase or a biologically active fragment of said DNA ligase, that is for the production of the DNA ligase or a biologically active fragment thereof in a suitable system, such as a host cell, driven by the genetic information contained in the nucleic acid, the nucleic acid may be provided in an expressible format, such as inserted or otherwise made a component of an expression cassette or vector, as is common in the art. Accordingly, aspects of the invention also relate to a nucleic acid encoding a DNA ligase or a biologically active fragment thereof as disclosed herein in an expressible format. Certain embodiments relate to an expression cassette or expression vector comprising the DNA ligase-encoding nucleic acids.

The term "expression cassette" encompasses a nucleic acid molecule, typically DNA, into which a coding sequence for a protein or proteins of interest may be inserted to be expressed, wherein said nucleic acid molecule comprises one or more nucleic acid sequence operably linked to and controlling the expression of the coding sequence (regulatory sequences), non-limiting examples of which include promoter sequences and transcription terminators. An "operably linkage" is a linkage in which regulatory sequences and sequences sought to be expressed are connected in such a way as to permit expression. For example, sequences, such as, e.g., a promoter and a coding sequence for a protein of interest, may be said to be operably linked if the nature of the linkage between said sequences does not: (1) result in the introduction of a frame-shift mutation, (2) interfere with the ability of the promoter to direct the transcription of the coding sequence, (3) interfere with the ability of the coding sequence to be transcribed from the promoter sequence. Hence, "operably linked" may mean incorporated into a genetic construct so that expression control sequences, such as a promoter, effectively control transcription / expression of a sequence of interest. The precise nature of transcriptional and translational regulatory sequences or elements required for expression may vary between expression environments, but typically transcriptional regulatory sequences include a promoter, optionally an enhancer, and a transcription terminator.

Reference to a "promoter" is to be taken in its broadest context and includes transcriptional regulatory sequences required for accurate transcription initiation and where applicable accurate spatial and/or temporal control of gene expression or its response to, e.g., internal or external (e.g., exogenous) stimuli. More particularly, "promoter" may depict a region on a nucleic acid molecule, preferably DNA

molecule, to which an RNA polymerase binds and initiates transcription. A promoter is preferably, but not necessarily, positioned upstream, i.e., 5', of the sequence the transcription of which it controls. Typically, in prokaryotes a promoter region may contain both the promoter per se and sequences which, when transcribed into RNA, will signal the initiation of protein synthesis (e.g., Shine-Dalgarno sequence). A promoter sequence can also include "enhancer regions", which are one or more regions of DNA that can be bound with proteins (namely the trans-acting factors) to enhance transcription levels of genes in a gene-cluster. The enhancer, while typically at the 5' end of a coding region, can also be separate from a promoter sequence, e.g., can be within an intronic region of a gene or 3' to the coding region of the gene.

In certain embodiments, promoters may be constitutive or inducible. A constitutive promoter is understood to be a promoter whose expression is constant under standard culturing conditions. Inducible promoters are promoters that are responsive to one or more induction cues. For example, an inducible promoter can be chemically regulated (e.g., a promoter whose transcriptional activity is regulated by the presence or absence of a chemical inducing agent such as an alcohol, tetracycline, a steroid, a metal, or other small molecule) or physically regulated (e.g., a promoter whose transcriptional activity is regulated by the presence or absence of a physical inducer such as light or high or low temperatures). An inducible promoter can also be indirectly regulated by one or more transcription factor that are themselves directly regulated by chemical or physical cues. Non-limiting examples of promoters include T7, U6, H1, retroviral Rous sarcoma virus (RSV) LTR promoter, the cytomegalovirus (CMV) promoter, the SV40 promoter, the dihydrofolate reductase promoter, the β-acting promoter, the phosphoglycerol kinase (PGK) promoter, and the EF1α promoter.

The terms "terminator" or "transcription terminator" refer generally to a sequence element at the end of a transcriptional unit which signals termination of transcription. For example, a terminator is usually positioned downstream of, i.e., 3' of a coding sequence encoding a polypeptide of interest. For instance, where a recombinant nucleic acid contains two or more coding sequences, e.g., successively ordered and forming together a multicistronic transcription unit, a transcription terminator may be advantageously positioned 3' to the most downstream coding sequence.

The terms "expression vector" or "vector" as used herein refer to nucleic acid molecules, typically DNA to which nucleic acid fragments may be inserted and cloned, i.e., propagated. Hence, a vector will typically contain one or more unique restriction sites, and may be capable of autonomous replication in a defined cell or vehicle organism such that the cloned sequence is reproducible. A vector may also preferably contain a selection marker, such as, e.g., an antibiotic resistance gene, to allow selection of recipient cells that contain the vector. Vectors may include, without limitation, plasmids, phagemids, bacteriophages, bacteriophage-derived vectors, PAC, BAC, linear nucleic acids, e.g., linear DNA,

transposons, viral vectors, etc., as appropriate (see, e.g., Sambrook et al., 1989; Ausubel 1992). Viral vectors may include *inter alia* retroviral vectors, lentiviral vectors, adenoviral vectors, or adeno-associated viral vectors, for example, vectors based on HIV, SV40, EBV, HSV or BPV. Expression vectors are generally configured to allow for and/or effect the expression of nucleic acids or open reading frames (ORF) introduced thereto in a desired expression system, e.g., *in vitro*, in a cell, organ and/or organism. For example, expression vectors may advantageously comprise suitable regulatory sequences.

Factors of importance in selecting a particular vector include inter alia: choice of recipient cell, ease with which recipient cells that contain the vector may be recognized and selected from those recipient cells which do not contain the vector, the number of copies of the vector which are desired in particular recipient cells, whether it is desired for the vector to integrate into the chromosome or to remain extra-chromosomal in the recipient cells, and whether it is desirable to be able to "shuttle" the vector between recipient cells of different species.

Expression vectors can be autonomous or integrative. A nucleic acid can be introduced into a cell in the form of an expression vector such as a plasmid, phage, transposon, cosmid or virus particle using genome engineering technologies known in the art. The recombinant nucleic acid can be maintained extrachromosomally or it can be integrated into the cell chromosomal DNA. Expression vectors can contain selection marker genes encoding protein required for cell viability under selected condition (e.g., URA3, which encodes an enzyme necessary for uracil biosynthesis, or LEU2, which encodes an enzyme required for leucine biosynthesis, or TRP1, which encodes an enzyme required for tryptophan biosynthesis, or PuroR, which encodes a puromycin resistance marker) to permit detection and/or selection of these cells transformed with the desired nucleic acids. Targeted nucleases can be used to introduce a nucleic acid in a cell. For example, the targeted nucleases include, but are not limited to, transcription activator-like effector nuclease (TALEN), zinc finger nuclease (ZNF), clustered regularly interspaced short palindromic repeats (CRISPR), CRISPR/Cas9, CRISPR/CPFL and combinations thereof.

Integrative vectors generally include a serially arranged sequence of at least a first insertable DNA fragment, a selectable marker gene, and a second insertable DNA fragment. The first and second insertable DNA fragments are each about 200 (e.g., about 250, about 300, about 350, about 400, about 450, about 500, or about 1000 or more) nucleotides in length and have nucleotide sequences which are homologous to portions of the genomic DNA of the cell species to be transformed. A nucleotide sequence containing a nucleic acid of interest for expression is inserted in this vector between the first and second insertable DNA fragments, whether before or after the marker gene. Integrative vectors can be linearized prior to transformation to facilitate the integration of the nucleotide sequence of interest into the cell genome. For example, CRISPR- or CRISPR/Cas9-based integrative vectors can be

used, such as wherein a Cas9 nuclease generates a DNA double strand break in the vector necessary to foster the insertion of the vector at the correct genomic site of the cell. As an example, a pAAVS1-puro-DNR vector can be used in combination with a pCas-Guide vector, wherein the DSB is generated by the Cas9 protein at the AVVS1 locus, followed by the insertion of the expression cassette at the genomic site.

Prior to introducing the vectors into a cell of interest, the vectors can be grown (e.g., amplified) in bacterial cells such as *Escherichia coli* (*E. coli*). The vector DNA can be isolated from bacterial cells by any of the methods known in the art, which result in the purification of vector DNA from the bacterial milieu. The purified vector DNA can be extracted extensively with phenol, chloroform, and ether, to ensure that no *E. coli* proteins are present in the plasmid DNA preparation, since these proteins can be toxic to mammalian cells.

In certain embodiments, the polynucleotide encoding the DNA ligase or biologically active fragment thereof as disclosed herein in an expressible format may be a DNA molecule configured to drive the expression of the DNA ligase or the biologically active fragment thereof (transcription of the DNA ligase coding sequence into corresponding DNA ligase or biologically active fragment thereof messenger RNA molecule and translation of the latter into a protein by cellular translation machinery) when the DNA molecule is introduced into an animal cell, preferably a non-bdelloid animal cell, such as an insect cell or a vertebrate cell, preferably a warm-blooded animal cell, even more preferably a mammalian cell. In certain embodiment, this may concern a DNA expression cassette or expression vector comprising a DNA coding sequence for the DNA ligase or biologically active fragment thereof as disclosed herein. In certain other embodiments, the nucleic acid encoding the DNA ligase or biologically active fragment thereof in an expressible format may be an RNA molecule, such as messenger RNA molecule (mRNA), configured to drive the expression of the DNA ligase protein or biologically active fragment thereof (translation of the coding sequence comprised by the RNA molecule into a protein by cellular translation machinery) when the mRNA molecule is introduced into an animal cell, preferably a non-bdelloid animal cell, such as an insect cell or a vertebrate cell, preferably a warm-blooded animal cell, even more preferably a mammalian cell. Such mRNA may be produced by any suitable means available in the art, such as by *in vitro* transcription from an expression cassette or vector comprising the DNA ligase or biologically active fragment thereof coding sequence.

There exist various well-known methods of introducing nucleic acids into animal cells, any of which may be used herein. At the simplest, the nucleic acid can be directly injected into the target cell. Other methods include fusion of the recipient cell with bacterial protoplasts containing the nucleic acid, the use of compositions like calcium chloride, rubidium chloride, lithium chloride, calcium phosphate,

DEAE dextran, cationic lipids or liposomes or methods like receptor-mediated endocytosis, biolistic particle bombardment ("gene gun" method), infection with viral vector (i.e., derived from lentivirus, adeno-associated virus (AAV), adenovirus, retrovirus or antiviruses), electroporation, and the like. Other techniques or methods which are suitable for delivering nucleic acid molecules to target cells include the continuous delivery of an nucleic acid molecule from poly(lactic-Co-Glycolic Acid) polymeric microspheres or the direct injection of protected (stabilized) nucleic acid molecule(s) into micropumps delivering the product. Another possibility is the use of implantable drug-releasing biodegradable microspheres. Also envisaged is encapsulation of nucleic acids in various types of liposomes (immunoliposomes, PEGylated (immune) liposomes), cationic lipids and polymers, nanoparticles or dendrimers, poly (lactic-Co-Glycolic Acid) polymeric microspheres, implantable drug-releasing biodegradable microspheres, etc.; and o-injection of nucleic acids with protective agent like the nuclease inhibitor aurintricarboxylic aicd. It shall be clear that also a combination of different above-mentioned delivery modes or methods may be used.

Further to the above illustration and guidance, the phrase "a cell genetically engineered to express" a protein of interest may thus in particular denote a cell altered or manipulated by man to comprise a nucleic acid encoding the protein of interest, such as a cell into which an exogenous nucleic acid encoding the protein of interest has been introduced or inserted by available technical means.

As already discussed, the present invention in particular envisages genetic engineering of cells to express the presently described DNA ligase protein or a biologically active fragment of said DNA ligase. The expression level of the DNA ligase or the biologically active fragment thereof, that is the quantity of the DNA ligase protein or biologically active fragment thereof produced by the cell, may vary within an acceptable range which may typically be determined by, on the lower end, the preference or necessity that the expression level is sufficient to endow the cell with sufficient DNA ligating activity, for example as evaluated by using an *in vitro* DNA ligation assay, and on the higher end, the preference or necessity to avoid reduction in the viability or functionality of the cells due to excessive overexpression of the exogenous DNA ligase protein or a biologically active fragment thereof. Such quantitative considerations can be applied by the skilled person.

Whereas the qualifier "isolated" need not be expressly recited with relation to the cells as intended herein, it may conveniently be included, as the present disclosure pertains to manipulation of cells outside of the body, *in vitro* or *ex vivo*, and subsequent therapeutic use of so-manipulated cells. The term "isolated" with reference to a particular component generally denotes that such component exists in separation from – for example, has been separated from or prepared and/or maintained in

separation from – one or more other components of its natural environment. More particularly, the term "isolated" as used herein in relation to a cell or cell population denotes that such cell or cell population does not form part of an animal or human body, for example the cell may be cultured, sorted or stored i*n vitro* or *ex vivo*.

The cell genetically engineered as disclosed herein is a non-bdelloid cell; preferably a non-bdelloid animal cell. In some embodiments, the cell genetically engineered as disclosed herein is an insect cell or a vertebrate cell. Preferably the cell is a warm-blooded animal cell, even more preferably, a mammalian cell. Mammalian cells are widely used in genetic engineering and for the production of recombinant proteins. In some embodiments, the mammalian cells can be mammalian cell lines, such as for example the Chinese Hamster Ovary (CHO) cells and Human Embryonic Kidney 293 (HEK293) cells. In some embodiments, mammalian cells can be primary cells, such as human T cells. In preferred embodiments, the cell is a human cell. In some embodiments, the cell is a rodent cell, such as a mouse cell, a rat cell, or a hamster cell.

In some embodiments, the cell genetically engineered as disclosed herein is an immune cell, preferably a T cell or an NK cell, more preferably a cytotoxic T cell. The term "T cell" or "T lymphocyte" as used herein is known in the art and is intended to include thymocyte, naïve T lymphocytes, immature T lymphocytes, mature T lymphocytes, resting T lymphocytes, or activated T lymphocytes. A T cell can be a T helper (Th) cell, for example a T helper 1 (Th1) or a T helper 2 (Th2) cell. The T cell can be a CD4+ T cell, a cytotoxic T cell (CD8+ T cell), a tumor infiltrating cytotoxic T cell (CD8+ T cell), CD4+CD8+ T cell, CD4-CD8- T cell, or any other subset of T cells. In preferred embodiments, the T cell is a cytotoxic T cell.

In some embodiments, the cell genetically engineered as disclosed herein is thus a NK cell. Native NK cells represent a distinct population of lymphocytes in terms of both phenotype and function. NK cells have a large granular lymphocyte morphology and express characteristic NK cell surface receptors, and lack both T cell receptor (TCR) rearrangement and T cell, B cell, monocyte and/or macrophage cell surface markers. The cells kill by releasing small cytoplasmic granules of proteins (perforin and granzyme) that cause the target cell to die by apoptosis. NK cells possess mechanisms distinguishing between potential "target" cells and healthy cells via a multitude of inhibitory and activating receptors that engage MHC class I molecules, MHC class I-like molecules, and molecules unrelated to MHC. NK cells can be detected by specific surface markers, such as CD16, CD56, and CD8 in humans with absence of CD3 expression.

In some further embodiments, the immune cell displays tumor specificity. For example, the immune cell has been isolated from a tumor of a subject. In preferred embodiments, the immune cell is a tumor infiltrating lymphocyte.

In some embodiments, the immune cell, such as the T cell or the NK cell, is further genetically engineered to express a chimeric antigen receptor (CAR) protein or a T cell receptor (TCR) protein. In preferred embodiments, the immune cell is further genetically engineered to express a tumor-specific CAR or TCR. CARs and TCRs are both genetically engineered receptors. These engineered receptors can be readily inserted into and expressed by immune cells, such as T cells or NK cells, in accordance with techniques known in the art. They can be used to recognize specific cells, such as tumor cells, activate the immune cell to attack and even destroy said tumor cells.

The term "Chimeric Antigen Receptor" or alternatively "CAR" refers to a recombinant polypeptide or a set of polypeptides, which, when expressed by an immune effector cell, endows the cell with specificity for a target antigen on the surface of a target cell, and with intracellular signal transduction. In some embodiments, a CAR comprises at least an extracellular antigen binding domain, a transmembrane domain and a cytoplasmic signaling domain comprising a functional signaling domain derived from a stimulatory molecule and/or a costimulatory molecule. The term "signaling domain" refers to the functional portion of a protein which acts by transmitting information within the cell to regulate cellular activity via defined signaling pathways by generating second messengers or functioning as effectors by responding to such messengers. Typically, a CAR may comprise a chimeric fusion protein, such that for example an antigen binding domain and an intracellular signaling domain are comprised within the same polypeptide chain. In other embodiments, the CAR may be formed by a set of polypeptides not contiguous with each other, such that for example an antigen binding domain and an intracellular signaling domain may be provided in separate polypeptide chains, configured to heterodimerize to form the CAR.

CAR constructs may typically be characterized as belonging to successive generations. In first-generation CARs the intracellular signaling domain contains or consists essentially of the zeta chain associated with the T cell receptor complex (CD3ζ) or the γ subunit of the immunoglobulin Fc receptor (FcRγ). The cytoplasmic signaling domain of second-generation CARs further comprise an intracellular costimulatory domain, i.e., a functional signaling domain derived from at least one costimulatory domain, i.e., a functional signaling domain derived from at least one costimulatory molecule, such as CD28, 4-1BB (CD137), DAP10, ICOS, or OX40 (CD134), and third-generation CARs include a combination of two or more such costimulatory endodomains.

In some embodiments, the immune cell is further genetically engineered to express a T cell receptor (TCR), preferably a tumor-specific TCR. The term "T cell receptor (TCR)" refers to a protein complex found on the surface of T cells that is responsible for recognizing fragments of antigens as peptides bound to major histocompatibility complex (MHC) molecules. TCRs are composed of two different subunits, α and β, and genetic engineering of TCRs to be used in immune therapy is generally intended to introduce new TCR α and β with selected peptide specificity.

The immune cells as disclosed herein, such as T cells or NK cells, can be isolated by any method known in the art. In certain embodiments, T cells or NK cells may be isolated from a subject, such as a human subject. In certain embodiments, T cells or NK cells may be isolated from cord blood or peripheral blood of a subject. In certain embodiments, T cells or NK cells may be *in vitro* differentiated from a hematopoietic stem cell or from an induced pluripotent stem (iPS) cell. In certain embodiments, T cells or NK cells may be derived from human peripheral blood mononuclear cells (PBMCs), unstimulated leukapheresis products (PBSC), human embryonic stem cells (hESCs), induced pluripotent stem cells (iPSCs), bone marrow, or umbilical cord blood. Methods to isolate or derive T cells or NK cells from such sources are well known in the art.

In certain embodiments, after isolation or production of the immune cells as disclosed herein, genetic engineering is performed to express a DNA ligase or a biologically active fragment thereof as disclosed herein in the immune cells. In certain embodiments, and as disclosed herein, the immune cell is also genetically engineered to express a CAR protein or a TCR protein, preferably a tumor-specific CAR or tumor-specific TCR.

In certain embodiments, the nucleic acid(s) introduced into the immune cells may be DNA expression cassette(s) or vector(s), from which mRNA (s) encoding the DNA ligase or biologically active fragment thereof and the CAR or TCR, optionally within a single ORF, can be transcribed using the cell's transcription machinery. In some embodiments, the DNA ligase or biologically active fragment thereof and the CAR or TCR are introduced into the immune cell using two separate DNA expression cassettes or vectors. In some other embodiments, the DNA ligase or biologically active fragment thereof and the CAR or TCR are introduced into the immune cell using a single expression cassette or vector. In other embodiments, the nucleic acid(s) introduced into the cells may be mRNA(s) encoding the DNA ligase or biologically active fragment thereof and the CAR or TCR, optionally within a single ORF, which can be translated by the cell's protein translation machinery. In certain preferred embodiments, the nucleic acid(s) such as mRNA(s) may be introduced into the starting population of immune cells by electroporation.

Optionally, the immune cells which comprise the nucleic acid(s) and which are capable of expressing the DNA ligase or biologically active fragment thereof and optionally the CAR or TCR may further be selected and/or expanded. Methods of selecting transfected or transduced cells are routine in the art and may be established positive selection markers, such as genes conferring resistance to blasticidin, geneticin, hygromycin B, puromycin or zeocin.

One aspect discloses the immune cell or immune cell population as disclosed herein for use as a medicament. Another aspect discloses a pharmaceutical composition comprising the immune cell or immune cell population as disclosed herein. In further embodiments, said cell, cell population or pharmaceutical composition are for use in a method of treating a proliferative disease, such as a tumor or cancer. For example, the cell, cell population or pharmaceutical compositions as disclosed herein are for use as immunotherapy in the treatment of a tumor or cancer. In particular embodiments, the cells, cell population or pharmaceutical compositions comprising said cells are for use as immunotherapy in the treatment of a tumor or cancer, wherein the cells are genetically engineered to express a DNA ligase or biologically active fragment thereof as taught herein. In more specific embodiments, the cells, cell population or pharmaceutical compositions comprising said cells are for use as immunotherapy wherein the cells are genetically engineered to a express a DNA ligase or biologically active fragment thereof as taught herein and to express a CAR or TCR, preferably a tumor-specific CAR or TCR as taught herein.

The term "immunotherapy" broadly encompasses any treatment that modulates an immune response, such as a humoral immune response, a cell-mediated immune response, or both. Immunotherapy comprises cell-based immunotherapy in which immune cells, such as T cells or NK cells, are transferred into the patient. The term also comprises an administration of substances or compositions, such as chemical compounds and/or biomolecules (e.g., antibodies, antigens, interleukins, cytokines, or combinations thereof), that modulate a subject's immune system. Examples of cancer immunotherapy include without limitation treatments employing monoclonal antibodies, for example Fc-engineered monoclonal antibodies against proteins expressed by tumor cells, immune checkpoint inhibitors, prophylactic or therapeutic cancer vaccines, adoptive cell therapy, and combinations thereof. Examples of immune checkpoint targets for inhibition include without limitation PD-1 (examples of PD-1 inhibitors include without limitation pembrolizumab, nivolumab, and combinations thereof), CTLA-4 (examples of CTLA-4 inhibitors include without limitation ipilimumab, tremelimumab, and combinations thereof), PD-L1 (examples of PD-L1 inhibitors include without limitation atezolizumab), LAG3, B7-H3 (CD276), B7-H4, TIM-3, BTLA, A2aR, killer cell immunoglobulin-like receptors (KIRs), IDO, and combinations thereof. In particular embodiments,

immunotherapy as used herein refers to cell-based immunotherapy, also known as adoptive cell therapy (ACT), which refers to the transfer of cells, most commonly immune-derived cells, such as in particular cytotoxic T cells (CTLs), back into the same patient or into a new recipient host with the goal of transferring the immunologic functionality and characteristics into the new host. If possible, use of autologous cells helps the recipient by minimizing tissue rejection and graft vs. host disease issues. Various strategies may for example be employed to genetically modify T cells by altering the specificity of the T cell receptor (TCR) for example by introducing new TCR α and β chains with selected peptide specificity. Alternatively, chimeric antigen receptors (CARs) may be used in order to generate immunoresponsive cells, such as T cells, specific for selected targets, such as malignant cells, with a wide variety of receptor chimera constructs having been described. Stem cell therapies in cancer commonly aim to replace bone marrow stem cells destroyed by radiation therapy and/or chemotherapy, and include without limitation autologous, syngeneic, or allogeneic stem cell transplantation. The stem cells, in particular hematopoietic stem cells, are typically obtained from bone marrow, peripheral blood or umbilical cord blood. Details of administration routes, doses, and treatment regimens of anti-cancer agents are known in the art, for example as described in "Cancer Clinical Pharmacology" (2005) ed. By Jan H. M. Schellens, Howard L. McLeod and David R. Newell, Oxford University Press. Active components of any combination therapy may be admixed or may be physically separated and may be administered simultaneously or sequentially in any order.

A further aspect provides a pharmaceutical composition comprising the engineered cells as taught herein and a pharmaceutically acceptable carrier.

The term "pharmaceutically acceptable" as used herein is consistent with the art and means compatible with the other ingredients of a pharmaceutical composition and not deleterious to the recipient thereof.

As used herein, "carrier" or "excipient" includes any and all solvents, diluents, buffers (e.g., neutral buffered saline or phosphate buffered saline), solubilizers, colloids, dispersion media, vehicles, fillers, chelating agents (e.g., EDTA or glutathione), amino acids (e.g., glycine), proteins, disintegrants, binders, lubricants, wetting agents, emulsifiers, sweeteners, colorants, flavourings, aromatisers, thickeners, agents for achieving a depot effect, coatings, antifungal agents, preservatives, stabilisers, antioxidants, tonicity controlling agents, absorption delaying agents, and the like. The use of such media and agents for pharmaceutical active substances is well known in the art. Such materials should be non-toxic and should not interfere with the activity of the cells.

The precise nature of the carrier or excipient or other material will depend on the route of administration. For example, the composition may be in the form of a parenteral acceptable aqueous solution, which is pyrogen-free and has suitable pH, isotonicity and stability. For general principles in medicinal formulation, the reader is referred to Cell Therapy: Stem Cell Transplantation, Gene Therapy, and Cellular Immunotherapy, by G. Morstyn & W. Sheridan eds., Cambridge University Press, 1996; and Hematopoietic Stem Cell Therapy, E. D. Ball, J. Lister & P. Law, Churchill Livingstone, 2000.

Liquid pharmaceutical compositions may generally include a liquid carrier such as water or a pharmaceutically acceptable aqueous solution. For example, physiological saline solution, tissue or cell culture media, dextrose or other saccharide solution or glycols such as ethylene glycol, propylene glycol or polyethylene glycol may be included.

The composition may include one or more cell protective molecules, cell regenerative molecules, growth factors, anti-apoptotic factors or factors that regulate gene expression in the cells. Such substances may render the cells independent of its environment.

Such pharmaceutical compositions may contain further components ensuring the viability of the cells therein. For example, the compositions may comprise a suitable buffer system (e.g., phosphate or carbonate buffer system) to achieve desirable pH, more usually near neutral pH, and may comprise sufficient salt to ensure isosmotic conditions for the cells to prevent osmotic stress. For example, suitable solution for these purposes may be phosphate-buffered saline (PBS), sodium chloride solution, Ringer's Injection or Lactated Ringer's Injection, as known in the art. Further, the composition may comprise a carrier protein, e.g., albumin (e.g., bovine or human albumin), which may increase the viability of the cells.

Further suitably pharmaceutically acceptable carriers or additives are well known to those skilled in the art and for instance may be selected from proteins such as collagen or gelatin, carbohydrates such as starch, polysaccharides, sugars (dextrose, glucose and sucrose), cellulose derivatives like sodium or calcium carboxymethylcellulose, hydroxypropyl cellulose or hydroxypropylmethyl cellulose, pre-gelatinized starches, pectin agar, carrageenan, clays, hydrophilic gums (acacia gum, guar gum, arabic gum and xanthan gum), alginic acid, alginates, hyaluronic acid, polyglycolic and polylactic acid, dextran, pectins, synthetic polymers such as water-soluble acrylic polymer or polyvinylpyrrolidone, proteoglycans, calcium phosphate and the like.

In an embodiment, the pharmaceutical cell preparation as defined above may be administered in a form of liquid composition. In embodiments, the cells or pharmaceutical composition comprising such can be administered systemically, topically, within an organ, at a site of organ dysfunction or lesion or at a site of tissue lesion.

Preferably, the pharmaceutical compositions may comprise a therapeutically effective amount of the desired cells. The term "therapeutically effective amount" refers to an amount which can elicit a biological or medicinal response in a tissue, system, animal or human that is being sought by a researcher, veterinarian, medical doctor or other clinician, and in particular can prevent or alleviate one or more of the local or systemic symptoms or features of a disease or condition being treated. Appropriate therapeutically effective amounts may be determined by a qualified physician with due regard to the nature of the desired cells, the disease condition and severity, and the age, size and condition of the subject.

Also provided are methods of producing said pharmaceutical compositions by admixing the cells of the invention with one or more additional components as described above as well as with one or more pharmaceutical excipients as described above.

Also disclosed is an arrangement or kit of parts comprising a surgical instrument or device for administration of the cells as taught herein or the pharmaceutical compositions as defined herein to a subject, such as for example systemically, for example, by injection, and further comprising the cells as taught herein or the pharmaceutical compositions as defined herein.

In an embodiment, the pharmaceutical composition as defined above may be administered in a form of a liquid composition.

The quantity of cells to be administered will vary for the subject being treated. In certain embodiments, the quantity of cells to be administered is between 10^2 to 10^{10} or between 10^2 to 10^9, or between 10^3 to 10^{10} or between 10^3 to 10^9, or between 10^4 to 10^{10} or between 10^4 to 10^9, such as between 10^4 and 10^8, or between 10^5 and 10^7, e.g., about $1x10^5$, about $5x10^5$, about $1x10^6$, about $5x10^6$, about $1x10^7$, about $5x10^7$, about $1x10^8$, about $5x10^8$, about $1x10^9$, about $5x10^9$, or about $1x10^{10}$ cells can be administered to a human subject. For example, such administration may be suitably distributed over one or more doses (e.g., distributed over 2, 3, 4, 5, 6, 7, 8 9 or 10 or more doses) administered over one or more days (e.g., over 1, 2, 3, 4 or 5 or more days). However, the precise determination of a therapeutically effective dose may be based on factors individual to each patient, including their size, age, tissue damage, and can be readily ascertained by those skilled in the art from this disclosure and the knowledge in the art. Suitably but without limitation, in a composition to be administered, cells may be present at a concentration between about 10^4/ml to about 10^9/ml, preferably between about 10^5/ml and about 10^8/ml, yet more preferably between about $1x10^6$/ml and about $1x10^8$/ml.

A further aspect provides the genetically engineered immune cells as taught herein for use as a medicament. Particularly provided are the engineered T cells or NK cells or the pharmaceutical

composition as taught herein for use in a method of treating a proliferative disease. Particularly provided are the genetically engineered immune cells or pharmaceutical composition as taught herein for use in a method of treating cancer. In further embodiments, said method of treating cancer using the genetically engineered cells or pharmaceutical composition as taught herein can be combined with radiation therapy, such as X-ray or carbon ion, and/or proton therapy, of the tumor or cancer.

Reference to "therapy" or "treatment" broadly encompasses both curative and preventative treatments, and the terms may particularly refer to the alleviation or measurable lessening of one or more symptoms or measurable markers of a pathological condition such as a disease or disorder. The terms encompass primary treatments as well as neo-adjuvant treatments, adjuvant treatments and adjunctive therapies. Measurable lessening includes any statistically significant decline in a measurable marker or symptom. Generally, the terms encompass both curative treatments and treatments directed to reduce symptoms and/or slow progression of the disease. The terms encompass both the therapeutic treatment of an already developed pathological condition, as well as prophylactic or preventative measures, wherein the aim is to prevent or lessen the chances of incidence of a pathological condition. In certain embodiments, the terms may relate to therapeutic treatments. In certain other embodiments, the terms may relate to preventive treatments. Treatment of a chronic pathological condition during the period of remission may also be deemed to constitute a therapeutic treatment. The term may encompass *ex vivo* or *in vivo* treatments as appropriate in the context of the present invention.

The terms "subject", "individual" or "patient" are used interchangeably throughout this specification, and typically and preferably denote humans, but may also encompass reference to non-human animals, preferably warm-blooded animals, even more preferably non-human mammals. Particularly preferred are human subjects including both genders and all age categories thereof. In other embodiments, the subject is an experimental animal or animal substitute as a disease model. The term does not denote a particular age or sex. Thus, adult and newborn subjects, as well as fetuses, whether male or female, are intended to be covered. The term subject is further intended to include transgenic non-human species; in particular transgenic rodents, such as mice, rats or hamsters; more in particular transgenic mice and/or transgenic rats.

The term "subject in need of treatment" or similar as used herein refers to subjects diagnosed with or having a disease as recited herein and/or those in whom said disease is to be prevented.

The term "neoplastic disease" generally refers to any disease or disorder characterized by neoplastic cell growth and proliferation, whether benign (not invading surrounding normal tissues, not forming metastases), pre-malignant (pre-cancerous), or malignant (invading adjacent tissues and capable of

producing metastases). The term neoplastic disease generally includes all transformed cells and tissues and all cancerous cells and tissues. Neoplastic diseases or disorders include, but are not limited to abnormal cell growth, benign tumors, premalignant or precancerous lesions, malignant tumors, and cancer. Examples of neoplastic diseases or disorders are benign, pre-malignant, or malignant neoplasms located in any tissue or organ, such as in the prostate, colon, abdomen, bone, breast, digestive system, liver, pancreas, peritoneum, endocrine glands (adrenal, parathyroid, pituitary, testicles, ovary, thymus, thyroid), eye, head and neck, nervous (central and peripheral), lymphatic system, pelvic, skin, soft tissue, spleen, thoracic, or urogenital tract.

As used herein, the terms "tumor" or "tumor tissue" refer to an abnormal mass of tissue that results from excessive cell division. A tumor or tumor tissue comprises tumor cells which are neoplastic cells with abnormal growth properties and no useful bodily function. Tumors, tumor tissue and tumor cells may be benign, pre-malignant or malignant, or may represent a lesion without any cancerous potential. A tumor or tumor tissue may also comprise tumor-associated non-tumor cells, e.g., vascular cells which form blood vessels to supply the tumor or tumor tissue. Non-tumor cells may be induced to replicate and develop by tumor cells, for example, the induction of angiogenesis in a tumor or tumor tissue.

As used herein, the term "cancer" refers to a malignant neoplasm characterized by deregulated or unregulated cell growth. The term "cancer" includes primary malignant cells or tumors (e.g., those whose cells have not migrated to sites in the subject's body other than the site of the original malignancy or tumor) and secondary malignant cells or tumors (e.g., those arising from metastasis, the migration of malignant cells or tumor cells to secondary sites that are different from the site of the original tumor). The term "metastatic" or "metastasis" generally refers to the spread of a cancer from one organ or tissue to another non-adjacent organ or tissue. The occurrence of the neoplastic disease in the other non-adjacent organ or tissue is referred to as metastasis.

Examples of cancer include but are not limited to carcinoma, lymphoma, blastoma, sarcoma, and leukemia or lymphoid malignancies. More particular examples of such cancers include without limitation: squamous cell cancer (e.g., epithelial squamous cell cancer), lung cancer including small-cell lung cancer, non-small cell lung cancer, adenocarcinoma of the lung, squamous carcinoma of the lung and large cell carcinoma of the lung, cancer of the peritoneum, hepatocellular cancer, gastric or stomach cancer including gastrointestinal cancer, pancreatic cancer, glioma, glioblastoma, cervical cancer, ovarian cancer, liver cancer, bladder cancer, hepatoma, breast cancer, colon cancer, rectal cancer, colorectal cancer, endometrial cancer or uterine carcinoma, salivary gland carcinoma, kidney or renal cancer, prostate cancer, vulvar cancer, thyroid cancer, hepatic carcinoma, anal carcinoma,

penile carcinoma, as well as CNS cancer, melanoma, head and neck cancer, bone cancer, bone marrow cancer, duodenum cancer, oesophageal cancer, thyroid cancer, or hematological cancer.

Other non-limiting examples of cancers or malignancies include, but are not limited to: Acute Childhood Lymphoblastic Leukemia, Acute Lymphoblastic Leukemia, Acute Lymphocytic Leukemia, Acute Myeloid Leukemia, Adrenocortical Carcinoma, Adult (Primary) Hepatocellular Cancer, Adult (Primary) Liver Cancer, Adult Acute Lymphocytic Leukemia, Adult Acute Myeloid Leukemia, Adult Hodgkin's Disease, Adult Hodgkin's Lymphoma, Adult Lymphocytic Leukemia, Adult Non-Hodgkin's Lymphoma, Adult Primary Liver Cancer, Adult Soft Tissue Sarcoma, AIDS-Related Lymphoma, AIDS-Related Malignancies, Anal Cancer, Astrocytoma, Bile Duct Cancer, Bladder Cancer, Bone Cancer, Brain Stem Glioma, Brain Tumors, Breast Cancer, Cancer of the Renal Pelvis and Urethra, Central Nervous System (Primary) Lymphoma, Central Nervous System Lymphoma, Cerebellar Astrocytoma, Cerebral Astrocytoma, Cervical Cancer, Childhood (Primary) Hepatocellular Cancer, Childhood (Primary) Liver Cancer, Childhood Acute Lymphoblastic Leukemia, Childhood Acute Myeloid Leukemia, Childhood Brain Stem Glioma, Glioblastoma, Childhood Cerebellar Astrocytoma, Childhood Cerebral Astrocytoma, Childhood Extracranial Germ Cell Tumors, Childhood Hodgkin's Disease, Childhood Hodgkin's Lymphoma, Childhood Hypothalamic and Visual Pathway Glioma, Childhood Lymphoblastic Leukemia, Childhood Medulloblastoma, Childhood Non-Hodgkin's Lymphoma, Childhood Pineal and Supratentorial Primitive Neuroectodermal Tumors, Childhood Primary Liver Cancer, Childhood Rhabdomyosarcoma, Childhood Soft Tissue Sarcoma, Childhood Visual Pathway and Hypothalamic Glioma, Chronic Lymphocytic Leukemia, Chronic Myelogenous Leukemia, Colon Cancer, Cutaneous T-Cell Lymphoma, Endocrine Pancreas Islet Cell Carcinoma, Endometrial Cancer, Ependymoma, Epithelial Cancer, Esophageal Cancer, Ewing's Sarcoma and Related Tumors, Exocrine Pancreatic Cancer, Extracranial Germ Cell Tumor, Extragonadal Germ Cell Tumor, Extrahepatic Bile Duct Cancer, Eye Cancer, Female Breast Cancer, Gallbladder Cancer, Gastric Cancer, Gastrointestinal Carcinoid Tumor, Gastrointestinal Tumors, Germ Cell Tumors, Gestational Trophoblastic Tumor, Hairy Cell Leukemia, Head and Neck Cancer, Hepatocellular Cancer, Hodgkin's Disease, Hodgkin's Lymphoma, Hypergammaglobulinemia, Hypopharyngeal Cancer, Intestinal Cancers, Intraocular Melanoma, Islet Cell Carcinoma, Islet Cell Pancreatic Cancer, Kaposi's Sarcoma, Kidney Cancer, Laryngeal Cancer, Lip and Oral Cavity Cancer, Liver Cancer, Lung Cancer, Lymphoproliferative Disorders, Macroglobulinemia, Male Breast Cancer, Malignant Mesothelioma, Malignant Thymoma, Medulloblastoma, Melanoma, Mesothelioma, Metastatic Occult Primary Squamous Neck Cancer, Metastatic Primary Squamous Neck Cancer, Metastatic Squamous Neck Cancer, Multiple Myeloma, Multiple Myeloma/Plasma Cell Neoplasm, Myelodysplastic Syndrome, Myelogenous Leukemia, Myeloid Leukemia, Myeloproliferative Disorders, Nasal Cavity and Paranasal Sinus Cancer, Nasopharyngeal Cancer,

Neuroblastoma, Non-Hodgkin's Lymphoma During Pregnancy, Non-melanoma Skin Cancer, Non-Small Cell Lung Cancer, Occult Primary Metastatic Squamous Neck Cancer, Oropharyngeal Cancer, Osteo-/Malignant Fibrous Sarcoma, Osteosarcoma/Malignant Fibrous Histiocytoma, Osteosarcoma/Malignant Fibrous Histiocytoma of Bone, Ovarian Epithelial Cancer, Ovarian Germ Cell Tumor, Ovarian Low Malignant Potential Tumor, Pancreatic Cancer, Paraproteinemias, Purpura, Parathyroid Cancer, Penile Cancer, Pheochromocytoma, Pituitary Tumor, Plasma Cell Neoplasm/Multiple Myeloma, Primary Central Nervous System Lymphoma, Primary Liver Cancer, Prostate Cancer, Rectal Cancer, Renal Cell Cancer, Renal Pelvis and Urethra Cancer, Retinoblastoma, Rhabdomyosarcoma, Salivary Gland Cancer, Sarcoidosis Sarcomas, Sezary Syndrome, Skin Cancer, Small Cell Lung Cancer, Small Intestine Cancer, Soft Tissue Sarcoma, Squamous Neck Cancer, Stomach Cancer, Supratentorial Primitive Neuroectodermal and Pineal Tumors, T-Cell Lymphoma, Testicular Cancer, Thymoma, Thyroid Cancer, Transitional Cell Cancer of the Renal Pelvis and Urethra, Transitional Renal Pelvis and Urethra Cancer, Trophoblastic Tumors, Urethra and Renal Pelvis Cell Cancer, Urethral Cancer, Uterine Cancer, Uterine Sarcoma, Vaginal Cancer, Visual Pathway and Hypothalamic Glioma, Vulvar Cancer, Waldenstrom's Macroglobulinemia, or Wilms' Tumor.

In certain embodiments, the cancer is a haematological malignancy. In certain embodiments, the cancer is a solid malignancy (solid tumor).

The terms "sample or "biological sample" as used throughout this specification include any biological specimen obtained (isolated, removed) from a subject. Samples may include without limitation organ tissue (e.g., primary or metastatic tumor tissue), whole blood, plasma, serum, whole blood cells, red blood cells, white blood cells (e.g., peripheral blood mononuclear cells), saliva, urine, stool (faeces), tears, sweat, sebum, nipple aspirate, ductal lavage, tumor exudates, synovial fluid, cerebrospinal fluid, lymph, fine needle aspirate, amniotic fluid, any other bodily fluid, exudate or secretory fluid, cell lysates, cellular secretion products, inflammation fluid, semen and vaginal secretions. Preferably, a sample may be readily obtainable by non-invasive or minimally invasive methods, such as blood collection ('liquid biopsy'), urine collection, faeces collection, tissue (e.g., tumor tissue) biopsy or fine-needle aspiration, allowing the provision / removal / isolation of the sample from a subject. The term "tissue" as used herein encompasses all types of cells of the human body including cells of organs but also including blood and other body fluids recited above. The tissue may be healthy or affected by pathological alterations, e.g., tumor tissue. The tissue may be from a living subject or may be cadaveric tissue. Particularly useful samples are those known to comprise, or expected or predicted to comprise, or known to potentially comprise, or expected or predicted to potentially comprise tumor cells and/or tumor microenvironment cells.

Any suitable weight or volume of a sample may be removed from a subject for analysis. Without limitation, a liquid sample may have a volume between 1 mL and 20 mL, e.g., 5 mL, 7.5 mL, 10 mL, 15 mL or 20 mL. A solid sample may have a weight of between 1 g and 20 g, e.g., 5 g, 7.5 g, 10 g, 15 g or 20 g.

In certain embodiments, the engineered immune cells as taught herein may be administered as the sole pharmaceutical agent (active pharmaceutical ingredient) or in combination with one or more other pharmaceutical agents where the combination causes no unacceptable adverse effects. For example, the immune cells may be combined with known anti-cancer therapy or therapies, such as for example surgery, radiotherapy, chemotherapy, biological therapy, or combinations thereof. The term "chemotherapy" as used herein is conceived broadly and generally encompasses treatments using chemical substances or compositions. Chemotherapeutic agents may typically display cytotoxic or cytostatic effects. In certain embodiments, the chemotherapeutic agent may be an alkylating agent, a cytotoxic compound, an anti-metabolite, a plant alkaloid, a terpenoid, a topoisomerase inhibitor, or a combination thereof. The term "biological therapy" as used herein is conceived broadly and generally encompasses treatments using biological substances or compositions, such as biomolecules, or biological agents, such as viruses or cells. In certain embodiments, a biomolecule may be a peptide, polypeptide, protein, nucleic acid, or a small molecule (such as primary metabolite, secondary metabolite, or natural product), or a combination thereof. Examples of suitable biomolecules include without limitation interleukins, cytokines, anti-cytokines, tumor necrosis factor (TNF), cytokine receptors, vaccines, interferons, enzymes, therapeutic antibodies, antibody fragments, antibody-like protein scaffolds, or combinations thereof. Examples of suitable biomolecules include but are not limited to aldesleukine, alemtuzumab, atezolizumab, bevacizumab, blinatumomab, brentuximab vedotine, catumaxomab, cetuximab, daratumumab, denileukin diftitox, denosumab, dinutuximab, elotuzumab, gemtuzumab ozogamicin, ^{90}Y-ibritumomab tiuxetan, idarucizumab, interferon A, ipilimumab, necitumumab, nivolumab, obinutuzumab, ofatumumab, olaratumab, panitumumab, pembrolizumab, ramucirumab, rituximab, tasonermin, ^{131}I-tositumomab, trastuzumab, Ado-trastuzumab emtansine, and combinations thereof. Examples of suitable oncolytic viruses include but are not limited to talimogene laherparepvec. Further categories of anti-cancer therapy include *inter alia* hormone therapy (endocrine therapy), immunotherapy, and stem cell therapy, which are commonly considered as subsumed within biological therapies. Examples of suitable hormone therapies include but are not limited to tamoxifen; aromatase inhibitors, such as atanastrozole, exemestane, letrozole, and combinations thereof; luteinizing hormone blockers such as goserelin, leuprorelin, triptorelin, and combinations thereof; anti-androgens, such as bicalutamide, cyproterone acetate, flutamide, and combinations thereof; gonadotrophin releasing hormone blockers, such as

degarelix; progesterone treatments, such as medroxyprogesterone acetate, megestrol, and combinations thereof; and combinations thereof.

In some embodiments, the genetically engineered cells or pharmaceutical compositions comprising said cells as disclosed herein are combined with radiation therapy of the tumor or cancer. The radiation therapy can be X-ray therapy, carbon ion therapy, proton therapy, electron beam radiotherapy or a combination thereof. The term "radiation therapy", also called "radiotherapy" is to be understood as any type of radiation that is used to treat cancer and that is intended to shrink and even destroy cancer cells. Radiation therapy can include external radiation therapy, internal radiation therapy and systemic radiation therapy. In external radiation therapy, external means are used wherein the energy beams come from a machine or system outside of the body. In contrast, internal radiation therapy, also known as brachytherapy, uses an internal implant, such as a tube or a wire, that is placed inside the body in or near the tumor site. Finally, systemic radiation therapy is to be understood as internal radiation therapy that is administered via the systemic route to the body.

In an aspect of the invention, the genetically engineered cells as taught herein can be used in bioproduction. The term "bioproduction" as used herein refers to laboratory- and industrial-scale production of a biological product (e.g., a biotherapeutic, or a vaccine) in a cell line, a cell lysate or a model *in vivo* platform (e.g., an egg). In preferred embodiments, the genetically engineered cells as taught herein are used for vaccine production.

In another aspect, a method of using a cell or cell population as disclosed herein in an environment characterized by radiation higher than terrestrial background radiation is disclosed. For example, said method can be a method of using the cell or cell population as disclosed herein in outer space, such as for example in deep space exploration.

Further provided is a method of using a transgenic animal as disclosed herein in an environment characterized by radiation higher than terrestrial background radiation, such as a method of using the transgenic animal as disclosed herein in outer space, such as in deep space exploration. In some embodiments, the method is not for treatment of the human or animal body by surgery or therapy. In some embodiments, the transgenic animal is a rodent; preferably a rodent selected from the group consisting of mice, rats, or hamsters; more preferably the transgenic animal is a mouse or a rat.

In some embodiments, the method of using a transgenic animal in an environment characterized by radiation higher than terrestrial background radiation, such as in outer space, includes a method for growing animals for agriculture.

It is apparent that there have been provided in accordance with the invention products, methods, and uses, that provide for substantial advantages as set forth above. While the invention has been described in conjunction with specific embodiments thereof, it is evident that many alternatives, modifications, and variations will be apparent to those skilled in the art in light of the foregoing description. Accordingly, it is intended to embrace all such alternatives, modifications, and variations as follows in the spirit and broad scope of the appended claims.

The above aspects and embodiments are further supported by the following non-limiting examples.

EXAMPLES

Methods

Bdelloid rotifer culture

All experiments were performed using isogenic *A. vaga* clones derived from a single individual from Meselson Laboratory. The cultures were maintained hydrated in cell flasks with natural spring water (Spa®), at 25° C and fed with lettuce filtrate.

Comparative proteomic analysis upon X-Ray irradiation of *A. vaga*

Each condition of irradiation was performed in duplicate with about 100,000 *A. vaga* individuals starved O/N in spring water including 1% of Penicillin/Streptomycin (ThermoFisher Scientific). X-ray radiations (1 kGy) were performed using a Precision X-Ray X-RAD 225 XL instrument (225 kV, 19 mA, filter 1, 30 cm). After radiation, the rotifers were incubated at 25°C until the extraction time. The proteins were extracted at t0h, t4h, t24h, and t72h post-irradiation. For protein extraction, rotifers were collected from the plate and centrifuged at 4,000 rpm, 4°C, 10 min. They were then washed twice in 10 ml of ice-cold PBS (centrifugation 4,000 rpm, 4°C, 10 min). After the last wash, the supernatant was removed and the pellet was resuspended in 300 µl of cold lysis buffer (10 mM Tris, pH 8.0; 1 mM EDTA, pH 8.0; 5 mM DTT; 1% v/v CHAPS; 1 M KCl; 1 mM PMSF, 25% v/v of glycerol and protease inhibitors). The resuspended pellet was then transferred to a cold Dounce and the rotifers were lysed by 1,000 Dounce with tight pestle B to crack the cuticle of the animals. The lysate was then recovered from the Dounce in an Eppendorf tube and the lysate was stocked O/N at -80°C. The cellular debris and insoluble proteins were then pelleted by ultracentrifugation at 42,000 rpm, 3h, 4°C. Supernatant containing soluble proteins was dialyzed against the dialysis buffer (20 mM Tris, pH 8.0; 20% glycerol; 100 mM KoAc; 0.5 mM EDTA, pH 8.0; 1 mM DTT) in Slide-A-Lyzer™ MINI Dialysis Device, 3.5K MWCO, 2 mL (ThermoFisher Scientific) according to the manufacturer protocol. After O/N dialysis at 4°C, the extracted proteins were recovered, the concentration was determined using Pierce 660nm

Protein Assay Reagent (ThermoFisher Scientific), and the proteins were flash frozen in liquid nitrogen before storage at -80°C. Filter-aided sample preparation (FASP) was used to generate tryptic peptides from protein mixtures prior to mass spectral analysis. The filter of Microcon-30kDa (Millipore) was first rinsed with 100 µl of 1% Formic acid (v/v) and centrifuged 15 min at 14,500 rpm. Fifty µg of protein extract were brought to 150 µl using Buffer UA (8 M urea in 0.1 M Tris/HCl pH 8.5). The extract was then added to the column and centrifuged 15 min at 14,500 rpm. The filter was then washed three times with 150 µl of Buffer UA (15 min at 14,500 rpm). For the reduction step, the samples were treated with 100 µl of DTT (8 mM DTT in UA buffer), incubated on a thermomixer at 24°C for 15 min, and centrifuged 10 min at 14,500 rpm. Filter was further rinsed with 100 µl of Buffer UA (15 min at 14,500 rpm). For the alkylation step, 100 µl of IAA (0.05 M iodoacetamide in UA) was added on the filter and incubated 20 min at 24°C in the dark before centrifugation 10 min at 14,500 rpm. Filter was then rinsed with 100 µl of Buffer UA (15 min at 14,500 rpm). The IAA was quenched by the addition of 100 µl of DTT. After an incubation of 15 min at 24°C, the column was centrifuged 10 min at 14,500 rpm, and further rinsed with 100 µl of Buffer UA. After centrifugation 10 min at 14,500 rpm, the samples were treated three times with 100 µl of Buffer ABC (0.05 M $NH4HCO_3$ in water) and centrifuged 10 min at 14,500 rpm. For the digestion step, 1 µg of trypsin (Promega) diluted in Buffer ABC was added to the column and incubated O/N at 24°C on a thermomixer under agitation (300 rpm). The digested proteins were recovered by placing the Microcon column on a protein LoBind tube of 1.5 ml followed by centrifugation 10 min at 14,500 rpm. The filter was rinsed with 40 µl of Buffer ABC and centrifuged 10 min at 14,500 rpm. The filtrate was then acidified with a solution of 10% v/v Trifluoroacetic acid (TFA) to reach 0.2% v/v TFA final concentration. The sample was finally dried in a speed-vac and the pellet was resuspended in 20 µl of 2% v/v acetonitrile (ACN) and 0.1% v/v TFA before mass spectrometry analysis.

The digested proteins were analyzed using nano-LC-ESI-MS/MS tims TOF Pro (Bruker) coupled with a UHPLC nanoElute (Bruker). Peptides were separated by nanoUHPLC (nanoElute, Bruker) on a 75 µm ID, 25 cm C18 column with integrated CaptiveSpray insert (Aurora) at a flow rate of 400 nl/min, at 50°C. LC mobile phase A was water with 0.1% formic acid (v/v) and B was ACN with formic acid 0.1% (v/v). The sample was loaded directly on the analytical column at a constant pressure of 800 bar. The digest (1 µl) was injected, and the organic content of the mobile phase was increased linearly from 2% B to 15% in 18 min, from 15% B to 25% in 9 min, from 25% B to 37% in 3 min and from 37% B to 95% in 5 min. Data acquisition on the tims TOF Pro was performed using Hystar 5.0 and otof-Control 6.0. tims TOF Pro data were acquired using 160 ms TIMS accumulation time, mobility (1/K0) range from 0.7 to 1.4 Vs/cm^2. Mass-spectrometric analysis were carried out using the parallel accumulation serial

fragmentation (PASEF) acquisition method. One MS spectra followed by six PASEF MSMS spectra per total cycle of 1.16 s. Two injections per sample were done.

Data analysis was performed using PEAKS Studio X+ with ion mobility module and Q module for label-free quantification (Bioinformatics Solutions Inc.). Protein identifications were conducted using PEAKS search engine with 15 ppm as parent mass error tolerance and 0.05 Da as fragment mass error tolerance. Carbamidomethylation was allowed as fixed modification, oxidation of methionine and acetylation (N-term) as variable modification. Enzyme specificity was set to trypsin, and the maximum number of missed cleavages per peptide was set at two. The peak lists were searched against the peptide database from *A. vaga* AV20 genome (Simion *et al.*, 2021; Sic Adv. 7(41): eabg4216). Peptide spectrum matches and protein identifications were normalized to less than 1.0% false discovery rate. Label-free quantitation (LFQ) method is based on expectation - maximization algorithm on the extracted ion chromatograms of the three most abundant unique peptides of a protein to calculate the area under the curve. For the quantitation, mass error and ion mobility tolerance were set respectively to 15 ppm and 0.08 1/k0. For the label-free quantitation results, peptide quality score was set to be ≥4 and protein significance score threshold was set to 20. The significance score is calculated as the -10log10 of the significance testing p-value (0.01). ANOVA was used as the significance testing method. Modified peptides were excluded and only proteins with at least two peptides were used for the quantitation. Total ion current was used to calculate the normalization factors.

Genomic integrity analysis

Genomic integrity was assessed by pulsed-field gel electrophoresis (PFGE) as described in Hespeels *et al.*, 2014 (J Evol Biol 27: 1334-1345) .

DNA ligase expression analysis by Western-blot

A custom polyclonal antibody targeting the copy B of the upregulated DNA ligase (AvLigE) was produced by a speedy 28-Day rabbit immunization program (Eurogentec). Two rabbits were injected with the synthetic peptide with sequence TTDNNEESNKRMKTD (SEQ ID NO: 9) located at the C-terminal end of the protein using KLH as carrier protein and wherein a carbonium ion (C+) has been added to the N-terminal end of the peptide. At the final bleed, the specific IgG targeting the DNA ligase peptide were purified by affinity.

For the Western-blot, 10 µg of protein extract (see above for preparation of the extract) were separated on a SDS-PAGE gel and then transferred to a PVDF membrane. The membrane was blocked for 1 hour with 5% w/v non-fat dry milk diluted in TBS-Tween 0.1% v/v. The anti-AvLigE-B antibody was then added (1:200) in TBS-T buffer supplemented with 1% w/v non-fat milk for 1h30 at RT.

Afterwards, the membrane was washed three times in TBS-T and incubated with 1:1,000 anti-rabbit secondary antibody coupled to HRP (ThermoFisher Scientific) in TBS-T - 1% milk for 1h at RT. The membrane was then washed three times in TBS-T and finally revealed with Super Signal West Pico Chemiluminescent substrate (ThermoFisher Scientific). The chemiluminescent signal was captured with an Imager AI600 (Amersham).

For tubulin detection, just after the revelation of the AvLigE-B signal, the same membrane was washed in TBS-T and incubated with 1:2,500 commercial anti-tubulin antibody (Sigma Aldrich) in TBS-T - 1% milk for 1h at RT. The membrane was then incubated with 1:2,500 anti-mouse secondary antibody coupled to HRP (ThermoFisher Scientific) in TBS-T - 1% milk for 1h at RT and revealed as described above.

Subcellular localization of the DNA ligase

The upregulated DNA ligase was localized by immuno-fluorescence within the bdelloid rotifer *A. vaga* following an irradiation at 1 kGy of X-Rays. The rotifers were first collected and concentrated by centrifugation (4,000 rpm, 4°C, 10 min) at different time points post-irradiation - t0h, t4h, t24h, t48h, and t72h. The individuals were resuspended in 20 µl of spring water and spread on a SuperFrost Plus microscope slide (ThermoFisher Scientific). The external cuticle of the rotifers was then removed by freeze-cracking. For that, the organisms were covered by a coverslip, gently squashed, and drown in liquid nitrogen. Immediately after the freezing, the coverslip was removed using a razor blade, and the slide was left at RT for few minutes. The rotifers were then fixed within 1% v/v Formaldehyde for 10 min at RT, followed by three washes in PBS. Blockage was realized by incubation of the slide within a solution of 1% w/v BSA and 0.3 % v/v Triton-X100 for 1 h at RT. The primary antibody (custom anti-AvLigE-B) was diluted at 1:150 in the blockage solution. A volume of 200 µl of that dilution was dropped on the slide, covered by a coverslip and incubated at 4°C O/N. The primary antibody was then washed three times in PBS. The secondary antibody - Alexa 568 Goat anti-rabbit (ThermoFisher Scientific) - was diluted at 1:150 in the blockage solution and a volume of 200 µl was dropped on the slide, and incubated for 1 h at RT. After 3 washes in PBS, the nuclei were stained with a 1:1,000 DAPI solution diluted in PBS and incubated 20 min at RT. After three final washes in PBS, the slides were mounted with Mowiol and sealed with a coverslip. The images were captured with a Zeiss LSM 900 with Airyscan2 in Z-stacks with constant parameters. Mounting and analysis of the images were done with the Fiji software (v2.1.0/1.53c).

DNA ligase gene evolutionary history analysis

To determine if the upregulated DNA ligase in *A. vaga* was acquired through a horizontal gene transfer (HGT), we used the Alienomics pipeline that detects both HGTs and contaminants in a genome

assembly (Simion *et al.*, 2021; Sci Adv. 7(41): eabg4216). Alienomics combines several genomic parameters such as gene taxonomy, GC content, sequencing depth, synteny and expression data, to detect HGTs.

For the phylogenetic analysis, the AvLigE copy B protein sequence (FUN-001822) (Simion *et al.*, 2021) was used as query to perform a BLAST search using default parameters against proteinic data from other bdelloid species for which genome assemblies were available (*i.e., Adineta ricciae, Rotaria magnacalcarata* and *Rotaria macrura*). Bedtools function getfasta was then used to extract the matched sequences into a fasta file. Sequences were then aligned using mafft with default parameters and the resulting alignment was inspected in order to manually fuse back non-overlapping hits from a given protein. De-aligned bdelloid sequences were kept for later use.

AvLigE-B sequence was also used to perform a diamond blastp search against uniref database using the following parameters: –more-sensitive -k 20000 –max-hsps 1 –evalue 1e-05 –query-cover 20. A total of 502 matched subject sequences were retrieved from the Uniref database, out of which only the best 150 hits were kept for subsequent analysis. These sequences were added to the bdelloid sequences previously retrieved as well as to sequences corresponding to ligases 1, 3, 4, B, C, D and K retrieved from NCBI.

The resulting 184 sequences were then aligned using mafft with default parameters. Alignment columns for which more than 80% missing data was observed were discarded using custom perl script, and the resulting reduced alignment was used as input for a phylogenetic analysis using iq-tree 1.6.1075 with the following parameters: -b 100 -m C20+R4. The resulting tree was analyzed with Archaeopterix with several lineages collapsed for clarity.

Conserved protein domains were predicted in the different sequences from DNA ligases using InterPro and illustrated with Affinity Designer. Only one representative sequence was illustrated when several sequences were collapsed. The kingdom classification for each organism was determined according to Ruggiero *et al.* (Plos One, 2015; 10: e0119248).

Quantitative PCR analysis

For the bdelloid rotifers *A. vaga* and *A. ricciae*, prior to irradiation, about 100,000 rotifers were starved 24h in the presence of 1% Penicillin/Streptomycin (ThermoFisher Scientific). Rotifers were irradiated at 1 kGy of X-Rays as described above.

For *M. verticillata*, fungi (ATCC 42662) were first grown on YM agar and then in liquid YM broth. The survival to X-Ray radiation doses of 0.05, 0.1, 0.2, 0.5 and 1.0 kGy was performed with a homogenous suspension of hyphae in liquid. After irradiation, an identical volume of liquid was plated on YM agar

and incubated at 25°C during 3 days. Survival potential was determined by doing a ratio of the diameter of the growing area in comparison to the non-irradiated condition. The experiment was performed in triplicate. The expression analysis of DNA ligase E from *M. verticillata* was performed after an irradiation of 0.2 kGy of X-Rays.

Total RNA of rotifers or fungi was extracted at different times post-irradiation: t0h, t4h, t24h, t48h, and t72h. A non-irradiated control was also added to the experiment. Total RNA extracts were prepared in Trizol (ThermoFisher Scientific) according to the manufacturer's instructions. Any residual DNA was digested with DNase I (ThermoFisher Scientific) and total RNA was reverse transcribed using the SuperScipt III First-Strand Synthesis System for RT-PCR (ThermoFisher Scientific). Gene expression level was determined with Applied Biosystems 7500 instrument (ThermoFisher Scientific) using the SsoAdvanced Universal SYBR Green Supermix (Biorad). The reactions were performed in triplicate for each condition.

To determine the relative abundance of the DNA ligase transcripts between irradiated and non-irradiated conditions, the real-time quantitative PCR experiments were performed with primers targeting the *gapdh* gene (= housekeeping gene) or the specific DNA ligase gene of each organism. All primer designs were performed using the online ePrimer3 software using the standard settings. The oligonucleotides used in this study are listed in Table 1. Relative gene expression data was determined using the Livak method (Livak et al., Methods; 2001; 25: 402-408).

Table 1. Oligonucleotide sequences used in the experiments (primer sequences, *in vitro* BER assay).

SEQ ID NO	Organism	Target gene	Accession number	Name	Sequence (5'>3')
10	*Adineta vaga*	*gapdh*	FUN_010333	Avaga_GAPDH_FW	TGTGCTGCAATCAAAGAAGC
11				Avaga_GAPDH_RV	CGACACGGTTTGAATAACCA
12		*DNA Ligase E_homoelog B*	FUN_001822	Avaga_LIGEb_FW1	CCACGATTTCCGACATTTCT
13				Avaga_LIGEb_RV1	CATCTTCGTCATCGGATGTG
14		*DNA Ligase E_homoelog A*	FUN_019185	Avaga_LIGEa_FW2	GCTGGTATTCCACGTTTTCC
15				Avaga_LIGEa_RV2	TCTTCATCGGATGCAACAAC
16		*DNA Ligase E_homoeolog B*	FUN_001822	Avaga_gblock_LIGEb	AGCTCATGACGCGATGTGTAGTGAAGCTGCTAAAGTTCTTCTTGCTGAAACATGGTCAGAAGATATTGA TCCGACAGGATGGTGGATGAGTGAAAAACTCGATGGTGTTCGAGCATACTGGAATGGAAAAAATTTTT ATTCAAGACAAGGAAATCTTTTTCATGCACCTGATTTTTTCAAGGCTGCACTACCTAAAGTTCCACTCGA TGGTGAAATATGGTGTGGCCGGGGTTTATTTCAAAAATGTGTTAGTATTGTCAAAAAACAAGCAAATA AAGTGGTGCCTCAAGATTATAAACTACTGACGTATCTAATTTTTGACGCACCAACTCAAGGAGGAAAAT ATGAAGATCGTGTCAAATGGTTACAAGCAAATATTCCACAAGATGATGATAACTGTTATGCTACAGTTG TTGGTATTAAAAAATGCGAAGGTTTAGCACAATTAAAACAATATTTAGCTGATGTGAACAAAGCTGGTG GAGAAGGTATTATGCTTCGTAAACCAGGTAGTCTTTATGAACATAAACGTTCAACTACTTTACTGAAAG TAAAAACGTTTTATGACGAAGAAGCTCATGTCATTGGTCATAAACCGAGTAAAAGCTTAATTGGTATGA CTGGTGCTCTTGAATGTGAATTACCGAATGGAAAACGATTTGATGTTGGTAGTGGACTTACTATGGATC AACGACGAAAACCACCAAAAATCGGTTCAGTCATTACATTTAAATTTCAAGAATTATCCAATAGTGGTA GCCCACGATTTCCGACATTTCTTCGTGTTCGTACTGATCTGACATGGAATGATGTTCTTGAAGCAGCTAA AACAAAAAAACCTGTCAGTGTCATACAGAAAGTGGTTCCATCGGCAAAATTATCTAAACAACATTCAAT ATTATTTTCTGTTATTCCATCACGTGATGCCAAAACCGATAAGAAAATCGTCACATCCGATGACGAAGAT GAAGATGCACCATCTACATCTACATCCACATCGACAAAGACGAATGATTCTCGACCGATGTGTAAATTT GGTGCAAAATGTTATCGAACAAATCCTGATCATCTAAAACAATATCAACATTCGTCATCCAGTACGAAA CCTGCCAAAACTACGGAAGCTGAATCGACCAATAAAACTGCTAAAACTGCAGAAGCAGAATCTACAAC CAAACCGCCCTGCACATTTGGTGCGAAATGTTATCGTACAAGTTCCATTCATTTAGCTACTTATTCTCATC CATCTAAAAATGGAACAAAGGAAGCGAGTGAAGAAGATATACTTGATACTCGTGAATTAGTTGAAGCT GAAGAACCTGCTTCACCAACCACAACTGATAATTTATTAGTCAAAGATAAAAATAAAGGCCTGACATTC GACGATGATGATGATGAAAGTAATGATCAAGAAATGGTTAGTCGTTCTAAAAAAGATTGGACCAATCT TGAAAATAAATGCAAAGAACAAAGCAAACGTTTAGCATATCTTGAAGAAATGTTTAAAACGACAAAAC GATCGAATTCAACAACTGATAATAATGAAGAAAGCAACAAACGTATGAAAACAGATTGAACGCGTGTC ATGAGCT

17		*DNA Ligase E_homoeolog A*	FUN_0191 85	Avaga_gblock_LIGE a	AGCTCATGACGCGATGAGTACAGCAGCAACACAGGTTCTTCTTGCAGAAACATGGAGCGAAGATATCG ATCCTACTGGTTGGTGGATGAGTGAAAAACTTGATGGTGTTCGAGCGTATTGGAGTGGGAGTAACTTT TATTCTCGACAAGGAAATCTTTTTTCACGTGCCTGATTTTTTTCAAGGCCGCACTTCCCAAAGTACCCTTGG ATGGAGAAATATGGTGTGGTCGTGGACTCTTTCAAAAATGTATAAGTATTATAAAGAAACAGGGAAAT AAAGTTGTTCCCGATGATTATAAGCTCTTGACTTATTTGATTTTCGATGCACCAACTCATGGAGGAAAAT ATGAAGATCGTGTCAAATGGTTACAGACGAATGTACCACAAGATGACGATAAATGTTATGCATCAGTT GTCGGTATAAAAAAAATGTCAAAATCTTGCAGATCTTAAACAATGTTTAGCTACTGTAAATGATGCCGGT GGAGAAGGAATTATGCTCAGGAAACCTGGTAGTCTATATGAAAATAAGCGTTCATCAACATTACTGAA AGTAAAAACGTTTTATGACGAAGAAGCTCTAGTTGTTGGTCATAAACCTGGAAAAGGAAATTGTACAG GAGTTTTAGGTGCTCTTCAATGTCAATTACCTAATGGAAAACGTTTCGATGTTGGAAGTGGCTTTGATA CAATATTATTCTCTATTATTCCATCACGTGATGTTAAAACTGGTAAGAAAGTTGTTGCATCCGATGAAGA AGGTGAACAAATATCTTCGTCATCACCTTCGACATCAACTTCAAAAACGAAGACAGATAACCAAGAAGT ATGTCAATATGGAGCGAAATGTTATCGAACTAATGCCGATCACCTCAAACAATATCAACATCCAACATC ATCTAAATTAACGAAATCGAAATCGAAAACAAAATCATCTCCTACCTCTACATTAAAAATTACACGACAA AGTTCTAAGAAATCAGAATCTCAAGATGAACCAACTTCACCAACTATCACAAGTAATCTACTTGTCAAA GATAAAAATAAAGGTCGTCATTTCGGTGACGATGATAAGGTAGATGATGATGATGATGATGAACC AGCTGCAACACCTATAACAAGTAATAAACGAAAACGAAATGTTGTTGAAAAAAAACCTGTTAGAAAAC GTACAAAGCAATCTTAGACGCGTGTCATGAGCT	
18		*DNA Ligase 1_homoeolog B*	FUN_0117 91	Avaga_LIG1b_FW	TGACGATGAATCCGATCAAA
19				Avaga_LIG1b_RV	TCGAATGGTGAAAACAACGA
20		*DNA Ligase 1_homoeolog A*	FUN_0117 93	Avaga_LIG1a_FW	CCATGTTAGCACATCCATCG
21				Avaga_LIG1a_RV	TGGTTGAAATGCATTCGGTA
22		*DNA Ligase 3_homoeolog A*	FUN_0079 98	Avaga_LIG3a_FW	TTTCTTCATGCTCGTGCAAC
23				Avaga_LIG3a_RV	TCGCGGATTTAGTTGCTTTT
24		*DNA Ligase 4_homoeolog A*	FUN_0237 39	Avaga_LIG4a_FW	CGTGCTGGTATGCTTTCTCA
25				Avaga_LIG4a_RV	TTCGACCATGGTTTCCATTT
26	*Adineta ricciae*	*gapdh*	g37006	Aricciae_GAPDH_FW	CTGGTATGGCTTTCCGTGTT

27				Aricciae_GAPDH_RV	GGCGTCGAAGATTGATGAGT
28		*DNA Ligase E_homoelog B*	g12513	Aricciae_LIGEb_FW	ATGCTATGCGACAGTCGTTG
29				Aricciae_LIGEb_RV	TGACCAATGACGAGTGCTTC
30		*DNA Ligase E_homoelog A*	g20670	Aricciae_LIGEa_FW	TGAAAGAGTCCTCGACCACA
31				Aricciae_LIGEa_RV	CTCGCAGAAACATGGAGTGA
32	*Mortierella verticillata*	*gapdh*	MVEG_05252	Mverticillata_GAPDH_FW	TGAACACCCAGTGCATTCAT
33				Mverticillata_GAPDH_RV	GGGCAAATATCGACAGGAAA
34		DNA Ligase E	MVEG_11954	Mverticillata_LIGE_FW	AAGCCCTTCAACCTGTTCCT
35				Mverticillata_LIGE_RV	TGGGATGGCTAAATTGCTTC
36	*Homo sapiens*	*gapdh*	*gapdh*	Hsapiens_GAPDH_FW	TGCACCACCAACTGCTTAG
37				Hsapiens_GAPDH_FW	GTTCAGCTCAGGGATGACC
38	n.a.	Oligos for *in vitro* BER assay (LIG03)	n.a.	n.a.	TCAACTCAGCAACTCCTT with a phosphorylation at the 5' end and an amino modifier group at the 3' end
39	n.a.	Oligos for *in vitro* BER assay (Loop02)	n.a.	n.a.	TTGGAGTTGCTGAGTTGATTCGTGAGCACCAACCGGTGCTCACGAA with a fluorescein at the 5' end

***In vitro* DNA ligation assay**

The *in vitro* DNA ligation activity of *A. vaga* cell extracts was assessed by using an assay described by Healing *et al.* (Nucleic Acids Res, 2019; 47: e61). The *A. vaga* whole protein extracts were prepared as described for the proteomic analysis except that the dialysis was performed in the absence of DTT to avoid the degradation of the antibody during the immuno-depletion. For the ligation assay with HEK 293T cell lines expressing AvLigE-A and AvLigE-B (see below), we prepared a nuclear protein extract using the NE-PER extraction kit (ThermoFisher Scientific) following the instructions from the manufacturer.

For the immuno-depletion of AvLigE, we used the Dynabeads Protein A Immunoprecipitation kit (ThermoFisher Scientific) following to the instructions of the manufacturer. A quantity of 5 µg of AvLigE-B antibody was bound and crosslinked to the magnetic beads using the BS3 crosslinker. A total of 100 µg of each protein extract was incubated in the presence of the beads during 1h at 4°C for the immuno-depletion. The immuno-depleted extracts were then recovered after capture of the beads by a magnet. A control experiment using nude magnetic beads was also performed. The protein concentration within all extracts was assessed using the Pierce 660nm Protein Assay Reagent (ThermoFisher Scientific) before the *in vitro* DNA ligation assay.

For the DNA ligation assay, we used the oligonucleotides LIG03 (SEQ ID NO: 38) and Loop02 (SEQ ID NO. 39) (Table 1), to create a complex structure containing a single-strand break. For the assay with protein extracts from *A. vaga*, the extracts were diluted in the dialysis buffer and used at the following final concentrations in the assay: 0.64 µg, 0.32 µg, 0.16 µg, 0.08 µg, 0.04 µg, 0.02 µg, and 0.01 µg. A negative control corresponding to the dialysis buffer without protein was also added to the analysis. For the assay with protein extracts from HEK 293T cell lines, the extracts were also diluted in the dialysis buffer and used at the following final concentrations in the assay: 1.28 µg, 0.64 µg, 0.32 µg, 0.16 µg, 0.08 µg, 0.04 µg, and 0.02 µg. A negative control corresponding to the dialysis buffer without protein was also added to the analysis. Each ligation assay was performed in technical quadruplicate following the procedure described in Healing *et al.* (2019) except that the reactions were performed for 2h at 15°C. Data were presented with Prism 7 as mean ± SD of the four technical replicates. Data were background corrected by subtraction of the mean value obtained for the controls performed with the dialysis buffer. Background corrected absorbance values were normalized to the maximum absorbance value obtained during the assay. Curve fit was performed from non-linear regression analysis using the Sigmoidal, 4PL, X is log (concentration) fitting tool. DNA ligation activities were compared by doing a ratio of the slopes obtained in the linear portion of the activity curves for each reaction.

Heterologous expression of AvLigE within the HEK 293T human cell line

AvLigE-A and AvLigE-B genes were stably transfected within the HEK 293T cell line by the CRISPR technology using the Stable cell line AVVS1 kit (Origene). HEK 293T cell line was maintained at 37°C and 5% CO_2 in Minimum Essential Medium Eagle (Sigma-Aldrich) supplemented with 10% of Fetal Bovine Serum and 1% of Penicillin/Streptomycin (ThermoFisher Scientific).

First, we inserted synthetic AvLigE-A and AvLigE-B genes (gblocks, IDT; Table 1) within the pAAVS1-puro-DNR plasmid at the *Sgf*I and *Mlu*I restriction sites using the Hifi DNA assembly cloning kit (New England Biolabs). This procedure cloned the genes of interest under the control of the constitutive CMV promoter. The whole cassette is bordered by two sequences homologous to the adeno-associated virus site 1 (AAVS1) locus on human chromosome 19, which is a safe insertion site within human cells. This safe harbor site is a transcriptionally active genomic region allowing robust and stable gene expression. This plasmid was co-transfected with the pCas-Guide vector following the instructions from Origene. Cas9 protein generates the DSB at the AVVS1 locus necessary to foster the insertion of the expression cassette at this genomic site. After several passages, the cell lines with stable insertions were selected by treatment with 0.5 µg/ml of puromycin. Several clones were selected and AvLigE-A/B insertions at AAVS1 locus were first verified by PCR using oligonucleotides amplifying the junction between the AAVS1 locus and the expression cassette (GE100033, Origene). Afterwards, we selected clones in which the mRNA expression of AvLigE-A and AvLigE-B was high and stable. To that purpose, total RNA from HEK 293T cell lines was first extracted using the Monarch total RNA Miniprep kit (New England Biolabs). The messenger RNAs corresponding to AvligE-A or AvLigE-B were then amplified and quantified by RT-qPCR using oligonucleotides described in Table 1 and the Luna Universal qPCR Master Mix (New England Biolabs). The relative abundance of these transcripts was reported to the quantity of the human *gapdh* house-keeping gene (Table 1). One clone expressing AvLigE-A or AvLigE-B which has passed these two quality control experiments was then selected for further experiments. A control cell line with an insertion of the empty pAAVS1-puro-DNR vector (HEK SCR) was also created.

To determine the proliferation potential of HEK 293T cell lines upon irradiation, we used the gold standard colony formation assay (CFA) which specifically test every cell in the population for its ability to undergo unlimited division. For this assay, HEK 293T cell lines (SCR, AvLigE-A, and AvLigE-B) were seeded in 6-well plates 24h before the irradiation experiments. X-ray radiations (0.002 and 0.004 kGy) were performed using a Precision X-Ray X-RAD 225 XL instrument (225 kV, 13.5 mA, filter 1, 50 cm). After the irradiation, the cells were trypsinized and diluted at different concentrations to get isolated colonies. Counting of cells was performed using the Moxi Z Mini automated cell counter (Orflo). After 13 days of growth, the medium was removed above the cells, rinsed carefully with PBS, and incubated

30 min with 2 ml of a mixture of 6.0% glutaraldehyde and 0.5% crystal violet. The fixation and staining mixture was rinsed with tap water and the colonies were counted in each conditions. Proliferation potential in each condition was measured by reporting the mean number of colonies growing at 0.002 and 0.004 kGy of X-Rays to the mean number of colonies growing in the non-irradiated control. The experiment was performed in two biological replicates, each including three technical replicates (n = 6). Statistical significance between the control cell line (HEK SCR) and the cell lines expressing AvLigE-A/B under the variable irradiation conditions was determined by multiple t-tests using the Holm-Sidak method, with alpha = 0.05 using Prism 7 software.

Results

DNA repair proteins are strongly upregulated upon irradiation of the bdelloid rotifer *A. vaga*, with a major upregulation of a DNA ligase

To get more insights into the DNA damage response (DDR) of *A. vaga*, we performed a comparative proteomic analysis upon radiation of hydrated individuals to 1 kGy of X-Rays. As shown through a pulsed field gel electrophoresis (PFGE), 1 kGy of X-Rays breaks the DNA in small fragments that are re-assembled in larger fragments post-irradiation suggesting an active DNA repair process (Figure1B). Whole-protein extracts of *A. vaga* rotifers (including a non-irradiated control) were prepared at different times post-irradiation (*i.e.*, t0h, t4h, t24h, and t72h), and the proteins were identified and quantified by a mass spectrometry analysis.

The results show that just at the end of radiation (t0h), there are no proteins up- or down-regulated in comparison to the non-irradiated control (Figure1A). This suggests that bdelloid rotifer *A. vaga* starts its repair processes upon release of the IR stress. However, from t4h post-irradiation, several proteins were significantly upregulated mainly including proteins involved in DNA repair such as a DNA ligase, DNA 3' Phosphatase, DNA polymerase *β*, and PARP-domain containing proteins (Figure1A). Interestingly, this set of proteins have functions that are required at different steps of the base excision repair (BER) pathway, confirming that base lesions are the major DNA damage occurring upon X-Ray radiation. Another protein, the alpha-protein kinase vwka, was also highly upregulated. This protein was described as involved in contraction processes in protozoa in collaboration with myosin II32 and is likely involved in muscle contraction of *A. vaga*, a function visually affected in irradiated bdelloid rotifers. All these proteins were consistently upregulated at t24h and t72h post-irradiation (Figure1A). Similar results of upregulation of these proteins were also observed after a radiation at 0.8 kGy of X-rays. At t72h, other protein families were also differentially expressed, which likely indicates a more general transcription recovery (Figure1A).

Among the upregulated proteins, the DNA ligase was particularly over-expressed at the different time points (up to 8.0 fold at t72h, Figure1A). We therefore further characterized this DNA ligase. The bdelloid rotifer *A. vaga* has a diploid genome structure with signatures of paleo-tetraploidy and around 40% of its genes has retained four copies, including this DNA ligase (Simion *et al.*, 2021). Two copies are located on homologous chromosomes 4a and 4b (copies A) and two are located on the homologous chromosomes 1a and 1b (copies B) (chromosomes 4 and 1 being homoeologs). Homoeologous copies A and B are divergent in sequence (68.5% of identity at the amino acid level) and only the B copies are upregulated upon irradiation, suggesting functional differences between homoeologous copies.

The expression pattern of this DNA ligase upon radiation was also verified by Western-blotting. A custom polyclonal antibody that targets the C-terminal peptide of copy B protein was produced. The result of the Western-blot shows that this protein is gradually expressed upon irradiation with the maximum expression at 72h (Figure 1C) following the same kinetics as the mass spectrometry analysis.

This antibody was also used to define the subcellular localization of the upregulated DNA ligase upon X-Ray irradiation of *A. vaga* by immuno-fluorescence. The result shows a specific gradation for the signal corresponding to the DNA ligase with a strong nuclear localization at t24h, which decreases at t48h and t72h (data not shown). This analysis suggests that the DNA ligase is mainly active during the first 24h following irradiation. This finding was further reinforced by the PFGE analysis that shows that the assembly of high molecular weight DNA fragments during the repair of DSBs mainly occurs during the first 24h and barely evolves between t24h and t72h (Figure 1B).

Altogether, the specific expression of this DNA ligase following desiccation or irradiation and its localization within somatic cell nuclei at time points at which the bdelloid rotifers exhibit signs of active DNA repair, strongly suggests that this DNA ligase protein is a key actor involved in DNA repair of *A. vaga*.

Upregulated DNA ligase has been horizontally-acquired by bdelloid rotifers

According to Hecox-Lea and Mark Welch, the identified DNA ligase has been acquired in bdelloid rotifers from a non-metazoan source by horizontal transfer. Using our recently developed Alienomics pipeline which aims to detect genes that have been horizontally acquired in assembled genomes (Simion et al., 2021 Sci Adv. 7(41): eabg4216), we confirmed the non-metazoan origin of the upregulated DNA ligase. Particularly, this gene was found to have a higher GC content than the average in *A. vaga* and the BLASTp analysis showed that the closest sequences were non-metazoans. To further identify the evolutionary history of this DNA ligase, we performed a phylogenetic analysis using the full-length copy B protein as query in the search of homologous sequences. The result shows

that both copies A and B of the DNA ligase of *A. vaga* are also found in all published genomes of bdelloid rotifers, and the two homoeolog sequences are well separated confirming their divergence (Figure 2). Surprisingly, the two *Rotaria* species, being desiccation-sensitive, also contain four copies of this DNA ligase suggesting that its acquisition occurred before the diversification of the bdelloid species and that this DNA ligase might not be correlated to their desiccation tolerance. This particular DNA ligase was not found in the monogonont rotifers, the sister group to the bdelloid clade, suggesting a specific acquisition within bdelloid rotifers.

The DNA ligase found in bdelloid rotifers is closely related to DNA ligases from non-metazoans including protozoa, chromista, fungi and bacteria (Figure 2). This protein contains two core functional domains, the ubiquitous ATP-dependent DNA ligase domain (pfam PF01068) and a DNA ligase OB-fold domain (pfam PF14743), that are more similar to the functional domains found in prokaryotic DNA ligases described as DNA ligase E. We therefore named this DNA ligase gene from *A. vaga* AvLigE-A and AvLigE-B for copy A and B, respectively. DNA ligase E forms a distinct evolutionary lineage from eukaryotic DNA ligases 1, 3 and 4 and prokaryotic DNA ligases B, C, and D35, 36 (Figure 2). All sequences of DNA ligases 1, 3, and 4 were found in *A. vaga* but only the AvLigE has been identified as horizontally acquired and upregulated in our proteomic analysis following IR. The main difference in the core ligase domains between AvLigE and the DNA ligases 1, 3 and 4 and DNA ligases B, C and D is the structure of the OB-fold domain (pfam PF04679, Figure 2), which likely reflects specific adaptation of this functional domain to fulfill specialized protein-DNA and protein-protein interactions during the DDR. DNA ligase K sequences found in kinetoplastids were added to the analysis because AvLigE was previously described as belonging to DNA ligase K (Hecox-Lea *et al*, 2018, BMC Evol Biol 18 (177)). The functional domains of the DNA ligase K are closer to the group of DNA ligase E, including the same core functional domains (pfam PF01068 and PF14743), which may suggest that DNA ligase K is also a derivative of DNA ligase E, remaining however phylogenetically more distant (Figure 2).

In addition to the core functional domains found in AvLigE, poly(ADP-ribose) (PAR)-binding zinc fingers (PBZ) domains are also present at the C-terminal end (Figure 2). AvLigE-B includes two PBZ domains whereas AvLigE-A contains only one PBZ domain (Figure 2). This structural variation might be related to a differential activity of the proteins, with copy B being the only one upregulated upon stress induction in *A. vaga*. The PBZ domains of both AvLigE-A and AvLigE-B are surrounded by proline-serine-threonine (PST)-rich repeats. Such PST-repeat regions have recently been shown to promote the chromatin interaction of the DNA repair protein MDC1 at the DNA DSBs in an H2AX-independent manner. MDC1 serves as a docking platform for other repair proteins at the site of the DNA damage. As H2AX was not found in *A. vaga*, proteins from bdelloids might include PST-repeats to specifically interact with chromatin and plausibly recruit other repair proteins at the site of DNA damage.

DNA ligase E is also upregulated in other bdelloid and fungus species upon irradiation

The phylogenetic analysis highlighted that the AvLigE protein had homologous sequences in other bdelloid rotifer species and in non-metazoans, including the fungus species *Mortierella verticillata* (Figure 2). To characterize the ubiquitous functional nature of this LigE upon irradiation, we measured the relative mRNA expression of this DNA ligase following radiation of the bdelloid rotifer *Adineta ricciae*, and the fungus *M. verticillata* (Figure 2). As *A. vaga* and *A. ricciae* can both tolerate high levels of irradiation, both species were exposed to 1 kGy of X-Ray and total RNA was extracted at different time points post-irradiation (t0h, t4h, t24h, t48h, and t72h) (Figure 3A-B). Because the radio-tolerance of *M. verticillata* was never reported in the literature, we first performed a survival assay that shows that, up to 0.2 kGy, survival of the fungi seems not affected, whereas there is a growth decrease at 0.5 and 1 kGy of X-ray radiation (Figure 3C). The expression level of the unique copy of DNA ligase E within *M. verticillata* was therefore measured after an exposure to 0.2 kGy of X-ray at the same time points as for the bdelloid rotifers (Figure 3C). The results of this experiment show that the DNA ligases E of all three organisms were specifically induced upon irradiation with a strong expression at t4h post-irradiation (Figure 3). For the bdelloid rotifers, only the copy B was upregulated as shown during the proteomic experiment of *A. vaga* (Figure 3A-B).

The proteomic analysis suggested that only AvLigE-B is upregulated upon irradiation. To further determine the functional importance of AvLigE-B in comparison to the other DNA ligases of *A. vaga* upon DNA damage induction, we measured their relative mRNA expressions at t4h following irradiation of *A. vaga* by 1 kGy of X-Ray (Figure 3D). The result shows that the AvLigE-B mRNA was induced up to 40 times higher than the other DNA ligases of *A. vaga*. Interestingly, there is more than 3-fold change in comparison to the AvLig3, which is the canonical DNA ligase involved in short-patch base excision repair (BER) and single strand break (SSB) repair (SSBR) (Figure 3D). This finding further highlights that AvLigE-B likely plays a major role during DNA damage repair in *A. vaga* and possibly all bdelloid rotifers.

AvLigE-B is a functional DNA ligase

Even though AvLigE-B is highly upregulated upon DNA damage induction within *A. vaga*, the exact function of this protein remains unknown. Particularly, the presence of this DNA ligase is striking as the bdelloid rotifers possess and express all canonical eukaryotic DNA ligases involved in DNA replication or repair (see Figure 3D). To obtain a functional characterization of this protein, we assessed the DNA ligation potential of AvLigE-B using an *in vitro* colorimetric ligation assay developed by Healing et al. (Nucleic Acids Res, 2019, 47 (e61)). The DNA ligation activity of a whole protein extract from irradiated and non-irradiated *A. vaga* was compared to the ligation activity of the same extract

depleted for AvLigE-B (Figure 4A). First, the Western-blot analysis did not detect any AvLigE-B in the non-irradiated protein extract which served as controls in the *in vitro* assay, while tubulin was detected (Figure 4B, lanes 1-4). On the contrary, upon irradiation, AvLigE-B production was upregulated and detected, while an incubation with an anti-AvLigE-B antibody clearly showed a depletion of this protein within the extract (Figure 4B, compare lanes 5 and 7). An elution of the proteins captured by the AvLigE-B antibody crosslinked to the magnetic beads confirmed the specificity of AvLigE-B capture, as the abundant tubulin protein, not recognized by the antibody, was only slightly captured by the beads (Figure 4B, lane 8). The incubation of the extract with nude magnetic beads did not deplete the AvLigE-B protein as expected (Figure 4B, lane 6).

The DNA ligation experiment is a microplate-based assay where increasing concentrations of the protein extracts are incubated in the presence of a complex structure between two oligonucleotides containing a SSB (Figure 4C). Repair of this SSB by ligation increases the concentration of fluorescein proportionally to the ligation activity, which is then measured by colorimetry (Figure 4C). The ligation activity of *A. vaga* protein extracts shows that irradiated individuals significantly enhance (> 5x) the global DNA ligation activity within the protein extract compared to a non-irradiated *A. vaga* extract (Figure 4D, compare IR to NI). This suggests that DNA ligation, and particularly DNA SSB repair, is a fundamental process activated by bdelloid rotifers upon irradiation. Depletion of AvLigE-B within the protein extract of irradiated *A. vaga* decreased the ligation activity by 1.6 fold, demonstrating that AvLigE-B is an active DNA ligase (Figure 4D, compare IR to IR-AvLigE-B depl.). However, ligation activity upon AvLigE-B depletion did not reach the ligation activity seen in the non-irradiated condition (Figure 4D, compare NI to IR-AvLigE-B depl.). This demonstrates that AvLigE-B is not the only ligation proficient enzyme upregulated upon irradiation, even though it contributes to about 35% of the ligation activity acquired following this stress.

Heterologous expression of AvLigE improves radio-tolerance of human cells

This study revealed that the horizontal acquisition of an atypical DNA ligase likely contributes to the high DNA damage tolerance of bdelloid rotifers. We hypothesized that the heterologous expression of AvLigE within human cells might also improve their DNA damage tolerance. For that purpose, the AvLigE-A and AvLigE-B genes were stably transfected and constitutively expressed within the HEK 293T human cell line (Figure 5A). An RT-qPCR experiment confirmed that both genes were expressed at a similar level and we then used the *in vitro* DNA ligation assay to define the ligation activity of AvLigE-A and AvLigE-B within human cells. AvLigE-B copy improved (i.e., 1.9 fold) the DNA ligation activity in comparison to the control HEK 293T cell line (HEK SCR), and surprisingly, the AvLigE-A copy even more significantly increased the *in vitro* ligation activity of the corresponding protein extract (> 5 fold, Figure 5B). Without wishing to be bound by any hypothesis, this finding might suggest that the AvLigE-B

protein is less stable within the human cells than the smaller AvLigE-A protein that could be efficiently folded and solubilized.

We then measured the survival potential of the different human cell lines by a colony formation assay upon 0.002 kGy and 0.004 kGy of X-Ray radiation. Upon 0.002 kGy of irradiation, only the cell line expressing the AvLigE-A protein could significantly improve the cell survival rate in comparison to the control cell line (Figure 5C). However, at 0.004 kGy of irradiation, both cell lines expressing AvLigE-A or AvLigE-B exhibited a significantly higher number of colonies than the control (Figure 5C). This increase in survival was following the *in vitro* DNA ligation activity of each cell line with a better survival for the cell line showing the highest DNA ligation activity (compare Figures 5B and 5C). This correlation suggests that the improved survival of these human cell lines upon irradiation is likely due to their enhanced DNA ligation activity. This improvement in proliferation potential of HEK 293T cells in particular upon expression of AvLigE-A was also confirmed by the morphological analysis of the colonies which exhibit very similar size and shape to the colonies formed in the non-irradiated condition (compare pictures of last column in Figure 5D). The morphology of the colonies developed by the control cell line (HEK SCR) and to some extent but less so by the cell line expressing AvLigE-B copy were strongly affected upon irradiation demonstrating a higher level of cell cycle arrest, necrosis, and apoptosis within these colonies (Figure 5D). Together, these results show that the expression of AvLigE-B copy and even more so AvLigE-A copy within human cells improves their DNA repair potential, leading to an increase in survival and proliferation rates following radiation exposure.

CLAIMS

1. An *in vitro* method for increasing tolerance of a cell to a stress factor, the method comprising genetically engineering the cell to express a DNA ligase that comprises an amino acid sequence with at least 70% sequence similarity to SEQ ID NO: 1 or SEQ ID NO: 2, or genetically engineering the cell to express a biologically active fragment of said DNA ligase, wherein the cell is a non-bdelloid cell.

2. The method of claim 1, wherein:

- the stress factor is an exogenous stress factor;
- the stress factor causes DNA damages in the cell;
- the stress factor causes DNA fragmentation in the cell;
- the stress factor causes generation of reactive oxygen species (ROS) in the cell;
- the stress factor is radiation;
- the stress factor is ionizing radiation;
- the stress factor is desiccation; or
- the stress factor is freezing.

3. The method of claim 1 or 2, wherein the tolerance of the cell to the stress factor is increased at least 1.05-fold, preferably at least 1.10-fold, such as at least 1.5-fold, at least 2.0-fold, or at least 2.5-fold, compared to a non-genetically engineered cell, as measured by cell survival fraction.

4. A cell genetically engineered to express a DNA ligase that comprises an amino acid sequence with at least 70% sequence similarity to SEQ ID NO: 1 or SEQ ID NO: 2, or genetically engineered to express a biologically active fragment of said DNA ligase; or a cell population comprising said cell, wherein the cell is a non-bdelloid cell.

5. A non-human transgenic animal genetically engineered to express a DNA ligase that comprises an amino acid sequence with at least 70% sequence similarity to SEQ ID NO: 1 or SEQ ID NO: 2, or genetically engineered to express a biologically active fragment of said DNA ligase, in at least some of its cells, tissues or organs.

6. A recombinant polynucleotide comprising a polynucleotide that encodes a DNA ligase that comprises an amino acid sequence with at least 70% sequence similarity to SEQ ID NO: 1 or SEQ ID NO: 2, or that encodes a biologically active fragment of said DNA ligase, wherein the polynucleotide that encodes the DNA ligase is operably linked to one or more heterologous regulatory sequences

configured to facilitate the expression of the DNA ligase or fragment thereof in a non-bdelloid cell or an animal; or a vector comprising said recombinant polynucleotide.

7. The method of anyone of claims 1 to 3 or the cell or cell population of claim 4 or the transgenic animal of claim 5 or the recombinant polynucleotide or vector of claim 6, wherein the non-bdelloid cell is an insect cell or a vertebrate cell, preferably a warm-blooded animal cell, more preferably a mammalian cell.

8. The method of anyone of claims 1 to 3 or the cell or cell population of claim 4 or the recombinant polynucleotide or vector of claim 6, wherein the cell is a human cell.

9. The method of anyone of claims 1-3 and 7-8 or the cell or cell population of any one of claims 4 and 7-8 or the transgenic animal of any one of claims 5 and 7-8 or the recombinant polynucleotide or vector of any one of claims 6-8, wherein:

- the DNA ligase comprises an amino acid sequence with at least 80% sequence similarity to SEQ ID NO: 1 or SEQ ID NO: 2;
- the DNA ligase comprises an amino acid sequence with at least 90% sequence similarity to SEQ ID NO: 1 or SEQ ID NO: 2;
- the DNA ligase comprises an amino acid sequence with at least 95% sequence similarity to SEQ ID NO: 1 or SEQ ID NO: 2;
- the DNA ligase comprises an amino acid sequence with at least 70% sequence identity to SEQ ID NO: 1 or SEQ ID NO: 2;
- the DNA ligase comprises an amino acid sequence with at least 80% sequence identity to SEQ ID NO: 1 or SEQ ID NO: 2; or
- the DNA ligase comprises an amino acid sequence with at least 90% sequence identity to SEQ ID NO: 1 or SEQ ID NO: 2.

10. The method of anyone of claims 1-3 and 7-8 or the cell or cell population of any one of claims 4 and 7-8 or the transgenic animal of any one of claims 5 and 7-8 or the recombinant polynucleotide or vector of any one of claims 6-8, wherein the DNA ligase comprises an amino acid sequence with at least 95% sequence identity to SEQ ID NO: 1 or SEQ ID NO: 2, such as wherein the DNA ligase comprises the amino acid sequence SEQ ID NO: 1 or SEQ ID NO: 2.

11. The method of anyone of claims 1-3 and 7-10 or the cell or cell population of any one of claims 4 and 7-10 or the transgenic animal of any one of claims 5 and 7-10 or the recombinant polynucleotide or vector of any one of claims 6-10, wherein the biologically active fragment comprises at least an ATP dependent DNA ligase domain and a DNA ligase OB-like domain.

12. The method of anyone of claims 1-3 and 7-11 or the cell or cell population of any one of claims 4 and 7-11 or the transgenic animal of any one of claims 5 and 7-11 or the recombinant polynucleotide or vector of any one of claims 6-11, wherein the DNA ligase is derived from Bdelloidea or bdelloid rotifers, preferably from Adinetidae, more preferably from *Adineta vaga*, preferably wherein the DNA ligase is an *Adineta vaga* Ligase E.

13. The method of anyone of claims 1-3 and 7-12 or the cell or cell population of any one of claims 4 and 7-12 or the recombinant polynucleotide or vector of any one of claims 6-12, wherein the cell is an immune cell, preferably a T cell or an NK cell, more preferably a cytotoxic T cell, optionally wherein:

- the immune cell displays tumor specificity, such as wherein the immune cell has been isolated from a tumor of a subject, preferably wherein the immune cell is a tumor infiltrating lymphocyte; or
- the immune cell is further genetically engineered to express a chimeric antigen receptor (CAR) protein or a T cell receptor (TCR) protein, preferably a tumor-specific CAR or TCR.

14. The cell or cell population of claim 13 for use as a medicament; or a pharmaceutical composition comprising the cell or cell population of claim 13.

15. The cell or cell population or pharmaceutical composition of claim 14 for use in a method of treating a proliferative disease, such as a tumor or cancer, such as immunotherapy of tumor or cancer, optionally wherein the treatment with the cell is combined with radiation therapy, such as X-ray, carbon ion, and/or proton therapy, of the tumor or cancer.

16. An *in vitro* method of using the cell or cell population of any one of claims 4 and 6-12 in an environment characterized by radiation higher than terrestrial background radiation, such as a method of using the cell or cell population of any one of claims 4 and 6-12 in outer space, such as in deep space exploration.

17. A method of using the transgenic animal of any one of claims 5 and 7-12 in an environment characterized by radiation higher than terrestrial background radiation, such as a method of using the transgenic animal of any one of claims 5 and 7-12, in outer space, such as in deep space exploration.

18. The method of claim 17, wherein the transgenic animal is a transgenic rodent; preferably a transgenic mouse.

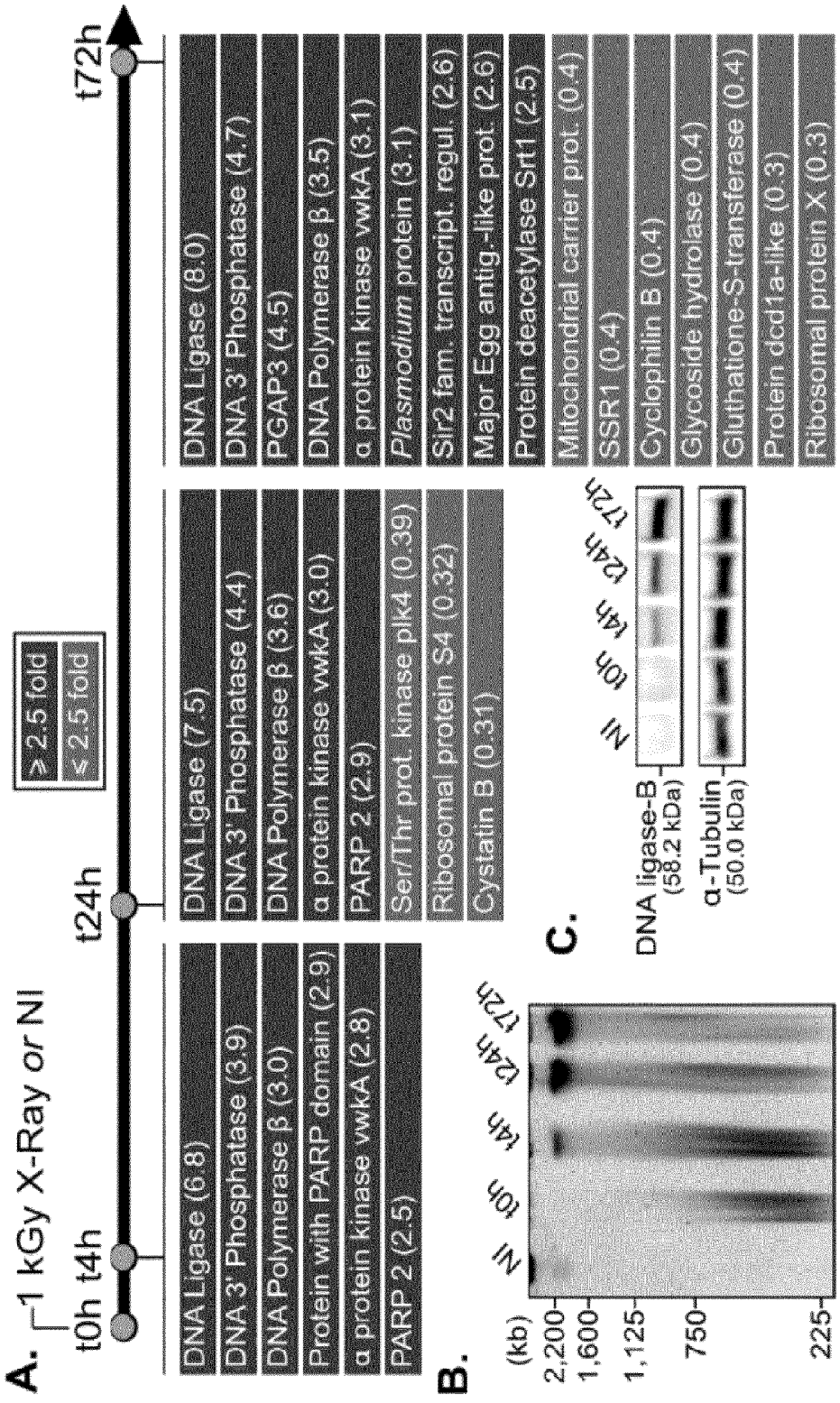
A.
1 kGy X-Ray or NI
t0h t4h
t24h
t72h
≥ 2.5 fold
≤ 2.5 fold
DNA Ligase (6.8)
DNA 3' Phosphatase (3.9)
DNA Polymerase β (3.0)
Protein with PARP domain (2.9)
α protein kinase vwkA (2.8)
PARP 2 (2.5)
DNA Ligase (7.5)
DNA 3' Phosphatase (4.4)
DNA Polymerase β (3.6)
α protein kinase vwkA (3.0)
PARP 2 (2.9)
Ser/Thr prot. kinase plk4 (0.39)
Ribosomal protein S4 (0.32)
Cystatin B (0.31)
DNA Ligase (8.0)
DNA 3' Phosphatase (4.7)
PGAP3 (4.5)
DNA Polymerase β (3.5)
α protein kinase vwkA (3.1)
Plasmodium protein (3.1)
Sir2 fam. transcript. regul. (2.6)
Major Egg antig.-like prot. (2.6)
Protein deacetylase Srt1 (2.5)
Mitochondrial carrier prot. (0.4)
SSR1 (0.4)
Cyclophilin B (0.4)
Glycoside hydrolase (0.4)
Glutathione-S-transferase (0.4)
Protein dcd1a-like (0.3)
Ribosomal protein X (0.3)
B.
NI
10h
14h
24h
72h
(kb)
2,200
1,600
1,125
750
225
C.
NI
10h
14h
24h
72h
DNA ligase-B
(58.2 kDa)
α-Tubulin
(50.0 kDa)

Figure 1

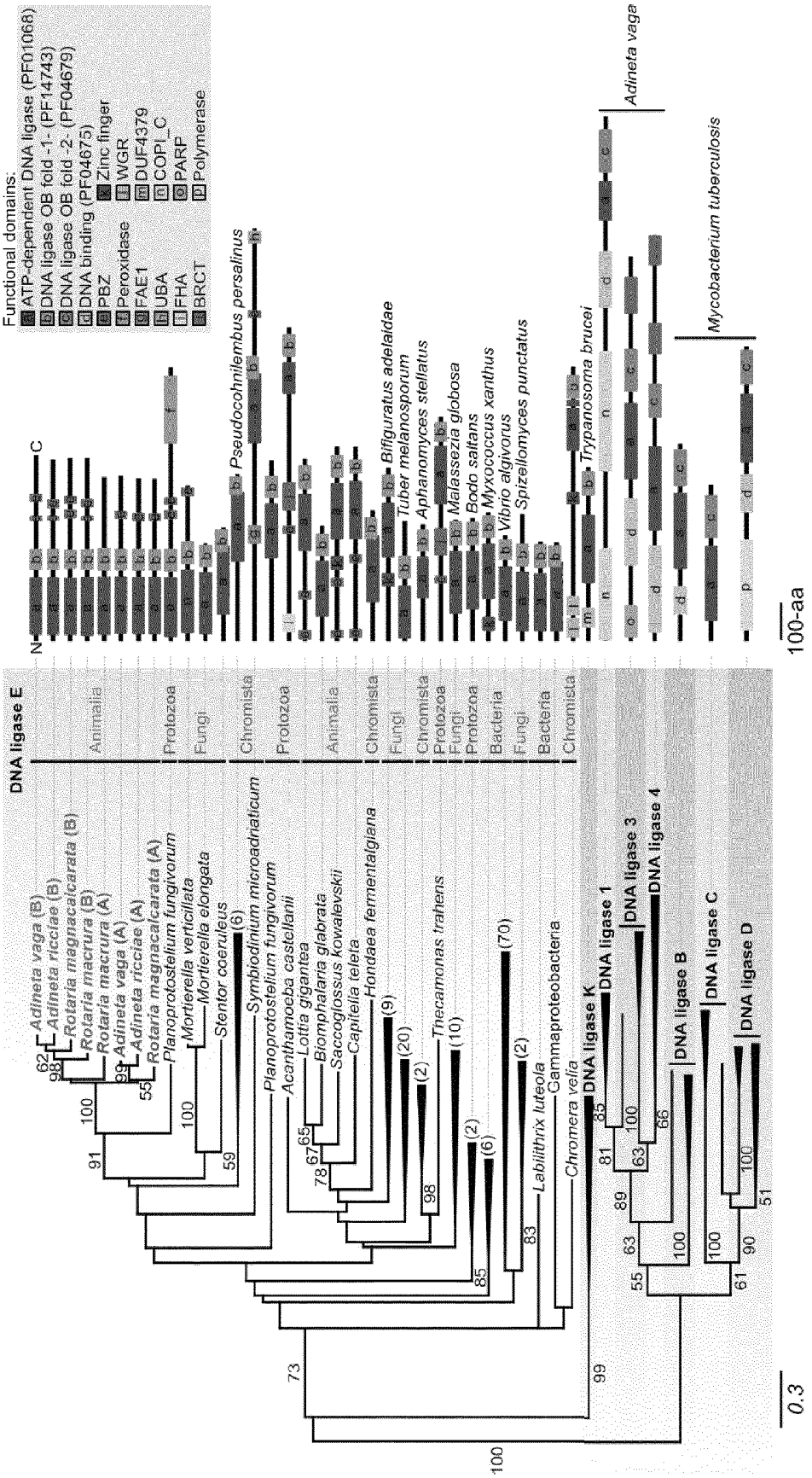
Functional domains:
ATP-dependent DNA ligase (PF01068)
DNA ligase OB fold -1- (PF14743)
DNA ligase OB fold -2- (PF04679)
DNA binding (PF04675)
PBZ
Zinc finger
Peroxidase
WGR
FAE1
DUF4379
UBA
COPI_C
FHA
PARP
BRCT
Polymerase
N
C
Pseudocohnilembus persalinus
Bifiguratus adelaidae
Tuber melanosporum
Aphanomyces stellatus
Malassezia globosa
Bodo saltans
Myxococcus xanthus
Vibrio algivorus
Spizellomyces punctatus
Trypanosoma brucei
Adineta vaga
Mycobacterium tuberculosis
100-aa
DNA ligase E
Adineta vaga (B)
Adineta ricciae (B)
Rotaria magnacalcarata (B)
Rotaria macrura (B)
Rotaria macrura (A)
Adineta vaga (A)
Adineta ricciae (A)
Rotaria magnacalcarata (A)
Planoprotostelium fungivorum
Mortierella verticillata
Mortierella elongata
Stentor coeruleus
Symbiodinium microadriaticum
Planoprotostelium fungivorum
Acanthamoeba castellanii
Lottia gigantea
Biomphalaria glabrata
Saccoglossus kowalevskii
Capitella teleta
Hondaea fermentalgiana
Thecamonas trahens
Labilithrix luteola
Gammaproteobacteria
Chromera velia
DNA ligase K
DNA ligase 1
DNA ligase 3
DNA ligase 4
DNA ligase B
DNA ligase C
DNA ligase D
Animalia
Protozoa
Fungi
Chromista
Protozoa
Animalia
Chromista
Fungi
Chromista
Protozoa
Fungi
Protozoa
Bacteria
Fungi
Bacteria
Chromista
0.3

FIGURE 2

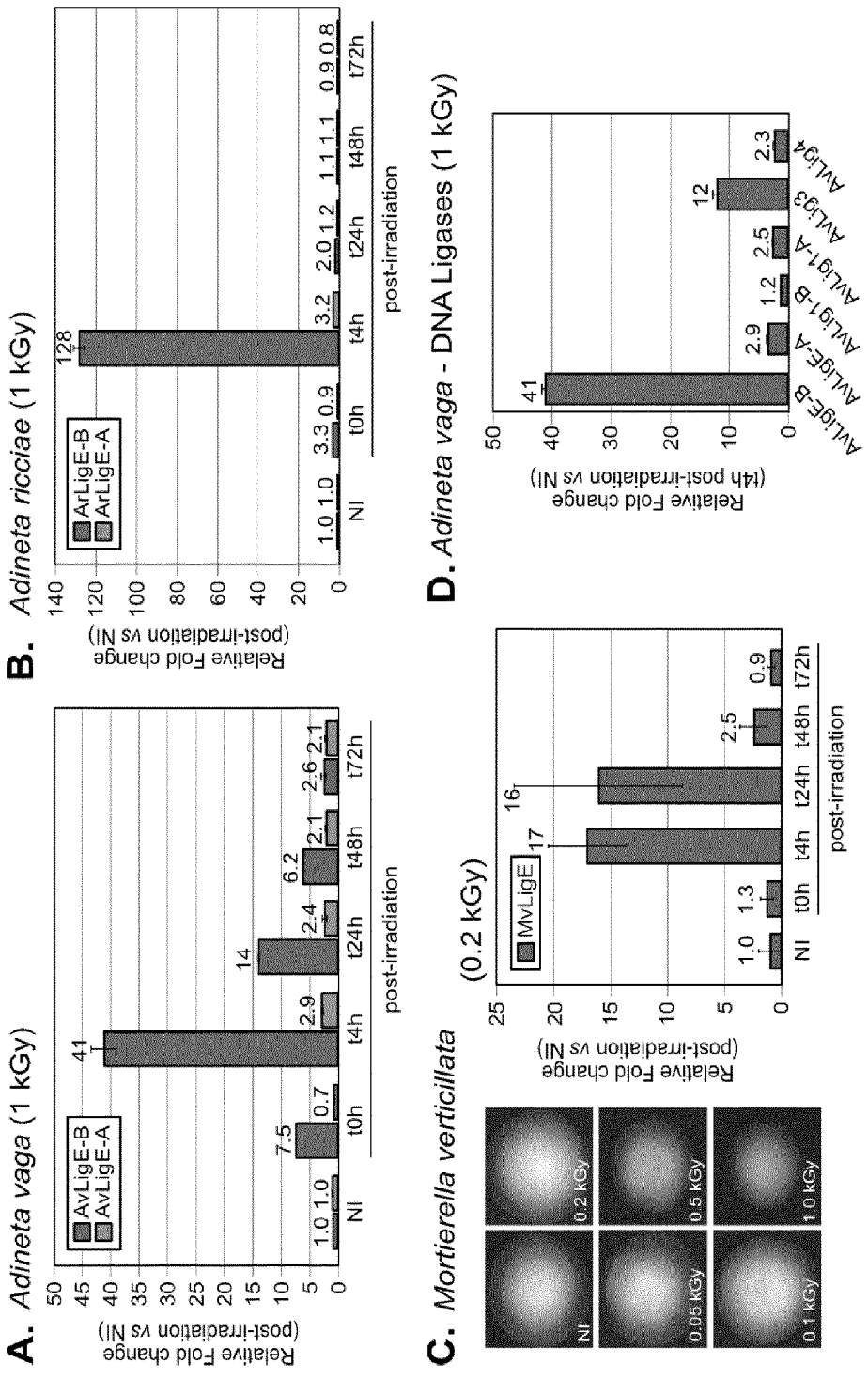
A. Adineta vaga (1 kGy)
AvLigE-B
AvLigE-A
Relative Fold change (post-irradiation vs NI)
NI
t0h
t4h
t24h
t48h
t72h
post-irradiation
B. Adineta ricciae (1 kGy)
ArLigE-B
ArLigE-A
C. Mortierella verticillata (0.2 kGy)
MvLigE
NI
0.05 kGy
0.1 kGy
0.2 kGy
0.5 kGy
1.0 kGy
D. Adineta vaga - DNA Ligases (1 kGy)
Relative Fold change (4h post-irradiation vs NI)
AvLigE-B
AvLigE-A
AvLig1-B
AvLig1-A
AvLig3
AvLig4

FIGURE 3

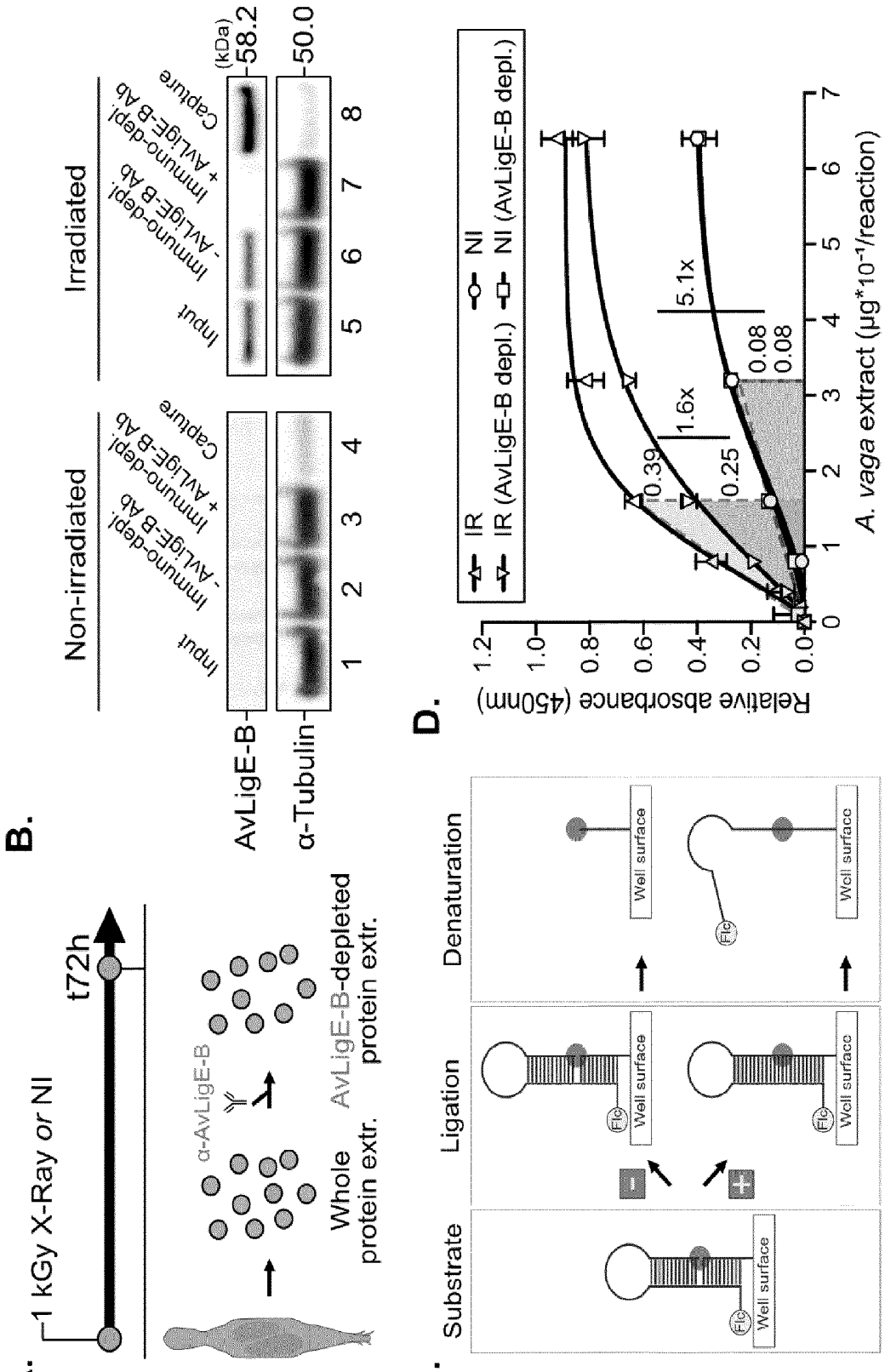
A.
1 kGy X-Ray or NI
t72h
α-AvLigE-B
Whole protein extr.
AvLigE-B-depleted protein extr.
B.
Non-irradiated
Irradiated
Input
Immuno-depl. -AvLigE-B Ab
Immuno-depl. +AvLigE-B Ab
Capture
(kDa)
AvLigE-B
α-Tubulin
58.2
50.0
1 2 3 4 5 6 7 8
C.
Substrate
Ligation
Denaturation
Well surface
Flc
D.
IR
IR (AvLigE-B depl.)
NI
NI (AvLigE-B depl.)
Relative absorbance (450nm)
A. vaga extract (µg*10⁻¹/reaction)
0.39
0.25
1.6x
5.1x
0.08
0.08

FIGURE 4

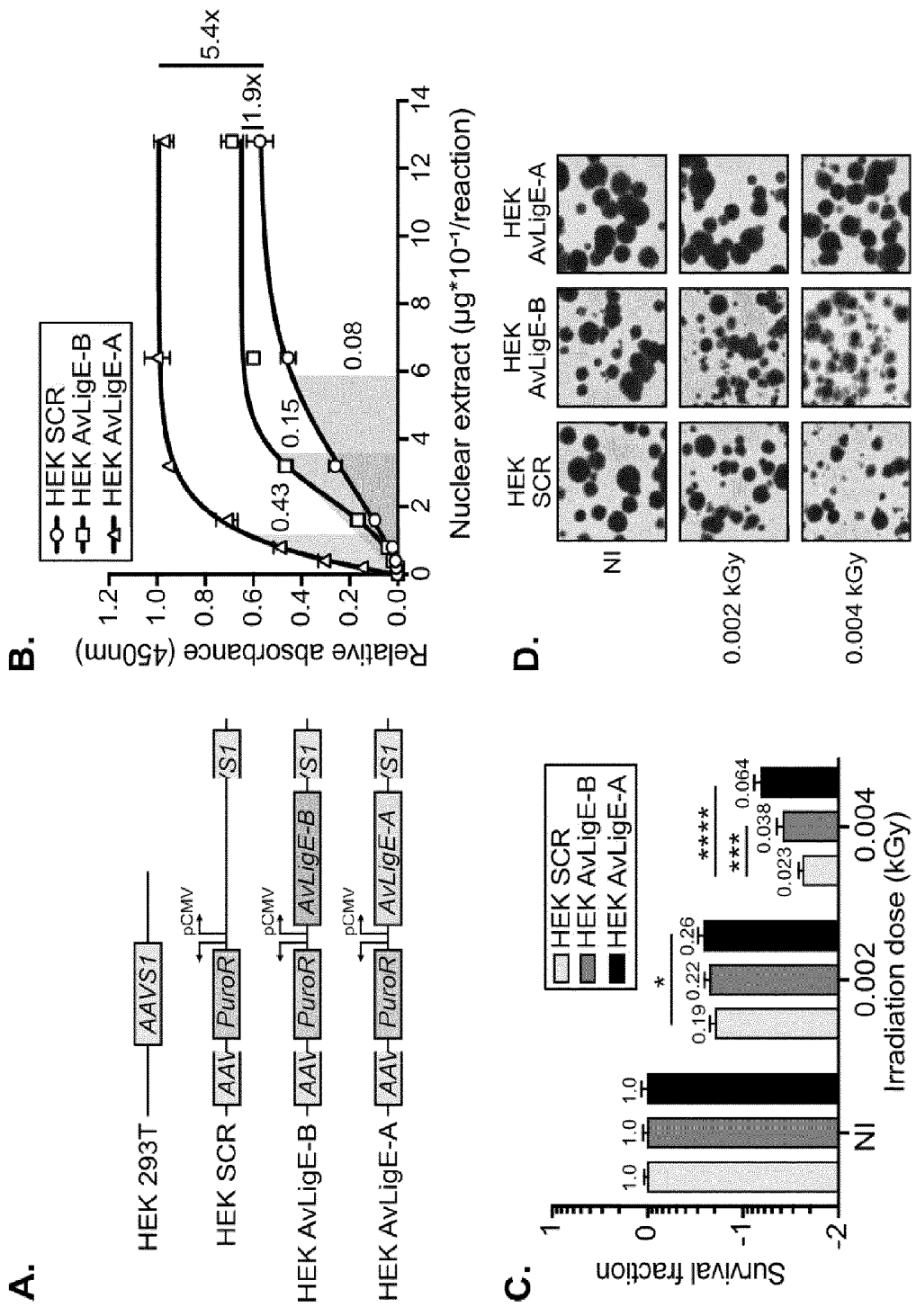
A.
HEK 293T
AAVS1
HEK SCR
AAV
pCMV
PuroR
'S1
HEK AvLigE-B
AvLigE-B
HEK AvLigE-A
AvLigE-A
B.
HEK SCR
HEK AvLigE-B
HEK AvLigE-A
Relative absorbance (450nm)
Nuclear extract (µg*10⁻¹/reaction)
0.43
0.15
0.08
1.9x
5.4x
C.
Survival fraction
Irradiation dose (kGy)
NI
0.002
0.004
1.0
0.19
0.22
0.26
0.023
0.038
0.064
*
***
****
D.
NI
0.002 kGy
0.004 kGy

FIGURE 5

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2023/054401

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N9/00 A61K38/53 C12N15/52
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, Sequence Search, BIOSIS, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	BETTEJ HECOX-LEA ET AL: "Evolutionary diversity and novelty of DNA repair genes in asexual Bdelloid rotifers", BMC EVOLUTIONARY BIOLOGY, BIOMED CENTRAL LTD, LONDON, UK, vol. 18, no. 1, 28 November 2018 (2018-11-28), pages 1-25, XP021263041, DOI: 10.1186/S12862-018-1288-9 abstract figures 2,5,8	1-18

-/--

[x] Further documents are listed in the continuation of Box C.

[ ] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance;; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance;; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
9 May 2023	17/05/2023

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Schmitz, Till

1

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Hecox-Lea: "Additional File 2 Evolutionary diversity and novelty of DNA repair genes in asexual Bdelloid rotifers", , 28 November 2018 (2018-11-28), XP055942261, Retrieved from the Internet: URL:https://static-content.springer.com/esm/art%3A10.1186%2Fs12862-018-1288-9/MediaObjects/12862_2018_1288_MOESM2_ESM.pdf [retrieved on 2022-07-13] page 13	1-18
Y	Anonymous: "UPI000A12E281 ¦ UniParc ¦ UniProt", , 13 November 2021 (2021-11-13), XP055942282, Retrieved from the Internet: URL:https://www.uniprot.org/uniparc/UPI000A12E281/entry [retrieved on 2022-07-13] sequence	1-18
Y	Anonymous: "UPI000A13B7B4 ¦ UniParc ¦ UniProt", , 13 November 2021 (2021-11-13), XP055942271, Retrieved from the Internet: URL:https://www.uniprot.org/uniparc/UPI000A13B7B4/entry [retrieved on 2022-07-13] sequence	1-18
X	DATABASE UniProt [Online] 29 September 2021 (2021-09-29), "SubName: Full=Hypothetical protein {ECO:0000313¦EMBL:CAF1328126.1};", XP002807076, retrieved from EBI accession no. UNIPROT:A0A815FNP9 Database accession no. A0A815FNP9 sequence	6

-/--

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE UniProt [Online] 29 September 2021 (2021-09-29), "SubName: Full=Hypothetical protein {ECO:0000313¦EMBL:CAF1147819.1};", XP002807077, retrieved from EBI accession no. UNIPROT:A0A814SQE7 Database accession no. A0A814SQE7 sequence -----	6
A	DATABASE UniProt [Online] 2 June 2021 (2021-06-02), "SubName: Full=Hypothetical protein {ECO:0000313¦EMBL:CAE0676137.1};", XP002807078, retrieved from EBI accession no. UNIPROT:A0A7S3Z9H8 Database accession no. A0A7S3Z9H8 sequence -----	1-18
A	LEE YOUNG HWAN ET AL: "Genome-wide identification of DNA double-strand break repair genes and transcriptional modulation in response to benzo[[alpha]]pyrene in the monogonont rotifer Brachionus spp", AQUATIC TOXICOLOGY, ELSEVIER, AMSTERDAM, NL, vol. 227, 28 August 2020 (2020-08-28), XP086262736, ISSN: 0166-445X, DOI: 10.1016/J.AQUATOX.2020.105614 [retrieved on 2020-08-28] the whole document -----	1-18
A	PAVLOPOULOU ATHANASIA ET AL: "Unraveling the mechanisms of extreme radioresistance in prokaryotes: Lessons from nature", MUTATION RESEARCH. REVIEWS IN MUTATION RESEARCH, ELSEVIER, AMSTERDAM, NL, vol. 767, 4 November 2015 (2015-11-04), pages 92-107, XP029484983, ISSN: 1383-5742, DOI: 10.1016/J.MRREV.2015.10.001 the whole document -----	1-18

1

INTERNATIONAL SEARCH REPORT

International application No.

PCT/EP2023/054401

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

   a. [X] forming part of the international application as filed.

   b. [ ] furnished subsequent to the international filing date for the purposes of international search (Rule 13*ter*.1(a)).

      [ ] accompanied by a statement to the effect that the sequence listing does not go beyond the disclosure in the international application as filed.

2. [ ] With regard to any nucleotide and/or amino acid sequence disclosed in the international application, this report has been established to the extent that a meaningful search could be carried out without a WIPO Standard ST.26 compliant sequence listing.

3. Additional comments: